

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя
(повне найменування вищого навчального закладу)
Факультет комп'ютерно-інформаційних систем і програмної інженерії
(назва факультету)
Кафедра комп'ютерних систем та мереж
(повна назва кафедри)

ПОЯСНЮВАЛЬНА ЗАПИСКА

до дипломного проекту (роботи)

магістра

(освітній ступінь (освітньо-кваліфікаційний рівень))

на тему: **Методи і засоби виявлення тональності текстової
інформації для вдосконалення комп'ютерних систем**

Виконав: студент (ка) 6 курсу, групи СІМ-61
спеціальності 123

«Комп'ютерна інженерія»

(шифр і назва спеціальності (напряму підготовки))

		<u>Гаван Р.Р.</u>
	(підпис)	(прізвище та ініціали)
Керівник		<u>Яцишин В.В.</u>
	(підпис)	(прізвище та ініціали)
Нормоконтроль		<u>Тиш Є.В.</u>
	(підпис)	(прізвище та ініціали)
Рецензент		<u>Баран І.О.</u>
	(підпис)	(прізвище та ініціали)

м. Тернопіль – 2019

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

(повне офіційне ім'я закладу вищої освіти)

Факультет Комп'ютерна інженерія систем і програмної інженерії

Кафедра Комп'ютерних систем та мереж

Освітній ступінь Магістр

Напрямок підготовки _____

Спеціальність 123 „Комп'ютерна інженерія“

(номер і назва)

(номер і назва)

ЗАТВЕРДЖУЮ

Завідувач кафедри КС

В.В. Шиншин
«30» листопада 2019 р.

ЗАВДАННЯ
НА ДИПЛОМНИЙ ПРОЕКТ (РОБОТУ) СТУДЕНТУ

Гавак Роман Русланович

(прізвище, ім'я, по батькові)

1. Тема проекту (роботи) Методи і засоби виявлення локальної текстової інформації для вдосконалення комп'ютерних систем

Керівник проекту (роботи) Шиншин В.В. доцент кафедри КС

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом по університету від «27» 09 2019 року №411/19

2. Термін подання студентом проекту (роботи) 23.12.2019

3. Вихідні дані до проекту (роботи) Матриця задачі інтегрований підхід до оптимізації локальної інформації, методи машинної навчання, матриці шукати оптимізовану локальну інформацію

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

1 Вступ 2. Аналіз сучасного стану досліджень в області аналізу локальної текстової інформації при використанні алгоритмів шукати 3. Побудова моделі та обчислення метрик оптимізації локальної інформації при використанні алгоритмів шукати 4. Система виявлення та аналізу локальної текстової інформації при використанні алгоритмів шукати 5. Оптимізація ефективності в шукати при та шукати в шукати 6. Результати аналізу шукати 7. Висновки

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

1. Формули і метри оптимізації 2. Об'єкти, результати, значення і методи оптимізації шукати шукати та шукати шукати результати 3. Формули шукати і шукати при виявленні та шукати локальної інформації 4. Матриця задачі інтегрований підхід 5. Метрики оптимізації локальної текстової інформації 6. Метрики F-MIRA 7. Шукати оптимізовану шукати комп'ютерів 8. USE CASE діаграма інженера та шукати шукати шукати 9. USE CASE діаграма шукати комп'ютерів та шукати шукати локальної інформації шукати шукати 10. Результати аналізу шукати шукати шукати 11. Результати аналізу шукати шукати шукати 12. Висновки

6. Консультанти розділів проекту (роботи)

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Обґрунтування економ. ефектів	Кирич Н.Б.		
Екологія	Косовина О.М. доц.		
Безпека в НС	Ступак В.С., м. викл. кадр. ОУ		
Оборонне протек.	Суховієва Т.М.	22.11.19	23.11.19

7. Дата видачі завдання 30.09.2019 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту (роботи)	Термін виконання етапів проекту (роботи)	Примітка
1	Аналіз умовного стану досліджень в області аналізу точності текстової інформації при вдосконаленні комп'ютерних систем	02.10.19	Виконав
2	Побудова моделей та обґрунтування методів і метрик оцінювання точності текстової інформації при вдосконаленні комп'ютерних систем	20.10.19	Виконав
3	Система виявлення та аналізу точності текстової інформації при вдосконаленні комп'ютерних систем	6.11.19	Виконав
4	Обґрунтування економічної ефективності	20.11.19	Виконав
5	Охорона протек. та безпека в надзвичайних ситуаціях	25.11.19	Виконав
6	Екологія	27.11.19	Виконав
	Попередній захист дипломної роботи	28.11.19	Виконав
	Захист дипломної роботи	23.12.19	

Студент Гаван Р.Р.
(підпис) (прізвище та ініціали)

Керівник проекту (роботи) Косовина О.М.
(підпис) (прізвище та ініціали)

АНОТАЦІЯ

Методи і засоби виявлення тональності текстової інформації для вдосконалення комп'ютерних систем // Дипломна робота// Гаван Роман Русланович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем та програмної інженерії, група СІм-61 // Тернопіль, 2019 // с. – 125, рис. – 41 , табл. – 21, аркушів А1 – 12 , додат. – 1, бібліогр. – 30.

КЛЮЧОВІ СЛОВА: ТОНАЛЬНІСТЬ, ТЕКСТ, ВИЯВЛЕННЯ, КОМП'ЮТЕРНА СИСТЕМА, ВДОСКОНАЛЕННЯ, МЕТОД, ЗАСІБ.

Метою роботи є дослідження методів і засобів визначення тональності текстової інформації у процесі супроводу комп'ютерних систем.

Задачі, які поставлено та розв'язано у дипломній роботі магістра полягають в наступному: аналіз наукових праць і практик реалізації методів і засобів виявлення та аналізу тональності текстової інформації; аналіз особливостей опрацювання відгуків користувачів при вдосконаленні комп'ютерних систем; обґрунтування інтелектуальних методів аналізу та визначення тональності у відгуках користувачів про комп'ютерні системи; розробка інтелектуального методу виявлення тональності відгуків користувачів; розробка програмного засобу збору, аналізу та виявлення тональності текстової інформації при супроводі комп'ютерних систем.

У першому розділі дипломної роботи досліджено джерела одержання інформації щодо властивостей комп'ютерних систем та виявлено, що для їх вдосконалення можуть використовуватись відгуки користувачів, які наявні у відкритому доступі та інциденти, що зберігаються у спеціалізованих комп'ютерних системах підтримки користувачів. Проведено аналіз методів комп'ютерної лінгвістики та прикладні аспекти їх застосування, що дало змогу визначити можливі шляхи впровадження методів машинного

навчання, препроцесингу текстової інформації на різних рівнях представлення тексту, а також застосування лінгвістично-інженерного підходу при побудові правил для опису шаблонів за частинами мовами.

У другому розділі дипломної роботи формалізовано представлення відгуків користувачів і задач, які при цьому необхідно розв'язати, у вигляді квантилю (сутність-аспект-тональність-користувач-час), що дало змогу врахувати позитивні і негативні властивості, реалізовані у комп'ютерних системах та в подальшому відобразити їх при вдосконаленні систем. Обгрунтовано необхідність застосування TF-IDF методів для виявлення та групування аспектів комп'ютерних систем, що позначаються на враженні кінцевих користувачів, обгрунтовано необхідність використання методів машинного навчання в комплексі з частотними методами синтаксичного аналізу при вирішенні задач вдосконалення комп'ютерних систем. Обгрунтовано застосування лінгвістично-інженерного підходу при оцінюванні тональності текстової інформації щодо вдосконалення комп'ютерних систем, що дає змогу врахувати якість аспектів на основі слів, що їх оцінюють та відстані між аспектами і такими словами. Встановлено та обгрунтовано доцільність застосування F-міри, що враховує повноту і точність оцінювання тональності відгуків користувачів і дає змогу виявити найбільш проблемні властивості комп'ютерної системи.

У третьому розділі, на основі UML діаграм, визначено функціональні вимоги та ролі користувачів при автоматизації процесу виявлення та оцінювання тональності текстової інформації, спроектовано схему бази даних для зберігання та маніпулювання даними щодо оцінювання тональності текстової інформації (відгуків користувачів), що дало змогу забезпечити гнучкість та масштабованість процесу вдосконалення комп'ютерних систем, а також спроектовано архітектуру системи виявлення та оцінювання тональності відгуків користувачів про комп'ютерну систему або її компоненти, що дало змогу реалізувати програмні модулі керування параметрами алгоритмів і моделей машинного навчання.

У четвертому розділі проведено розрахунки економічної доцільності проведення науково-дослідної роботи і визначено техніко-економічні показники щодо необхідності проведення таких робіт.

У дипломній роботі також розглянуто питання охорони праці та безпеки в надзвичайних ситуаціях, екології.

ABSTRACT

Methods and tools of text information tonality detection aimed at computer systems improvement // Master thesis// Havan Roman Ruslanovych// Ternopil Ivan Pul'uj National Technical University, Faculty of Computer Information Systems and software engineering, group CIm-61 // Ternopil, 2019// p. - 125, fig. – 41, table. – 21, Sheets A1 - 12, Add. - 1, Ref. - 30.

KEY WORDS: SENTIMENT, TEXT, EXTRACTION, COMPUTER SYSTEM, IMPROVING, METHOD, TOOL

The purpose of this master's thesis is to investigate the methods and means of determining the sentiment of text information in the process of computer systems support.

The tasks set and solved in the master's thesis are as follows: analysis of scientific works and practices of methods and means of detection and analysis of the sentiment of text information; analysis of the features of processing user feedback when improving computer systems; substantiation of intelligent methods of analysis and determination of the sentiment in user feedback about computer systems; developing an intelligent method for detecting the sentiment of user feedback; development of software for collecting, analyzing and detecting the sentiment of text information in the support of computer systems.

The first chapter of the thesis examines the sources of information about the properties of computer systems and found that they can be used to improve feedback from users who are in the public domain and incidents stored in specialized computer support systems. The analysis of computer linguistics methods and the applied aspects of their application were carried out, which made it possible to identify possible ways of introducing machine learning methods, reprocessing textual information at different levels of text presentation, as well as

applying a linguistic-engineering approach in constructing rules for describing patterns by parts of languages.

The second chapter of the thesis formalized the presentation of user reviews and tasks that need to be solved in the form of a quantile (essence-aspect-tone-user-time), which allowed to take into account the positive and negative properties realized in computer systems and further reflect them as systems improve. The necessity of using TF-IDF methods for identifying and grouping aspects of computer systems that affect the end-user experience is substantiated, the necessity of using machine learning methods in combination with frequency parsing methods in solving problems of improving computer systems is substantiated. The use of a linguistic-engineering approach in evaluating the tonality of textual information for the improvement of computer systems is justified. The expediency of applying the F-measure, which takes into account the completeness and accuracy of the evaluation of the tone of user feedback and allows to identify the most problematic properties of the computer system, has been established and substantiated.

The third chapter, based on UML diagrams, defines the functional requirements and roles of users in automating the process of detecting and evaluating the tone of text information, designed a database schema for storing and manipulating the data for assessing the tone of text information (user feedback), which allowed for flexibility and scalability of the process of improving computer systems, as well as designed the architecture of the system to detect and evaluate the tone of user feedback on computer system or its components, which made it possible to implement software modules for controlling the parameters of algorithms and machine learning models.

In the fourth chapter, the calculations of the economic feasibility of conducting research work are carried out and the technical and economic indicators for the necessity of carrying out such works are determined.

The diploma thesis also deals with the issues of occupational safety and health, ecology.

ЗМІСТ

ПЕРЕЛІК ОСНОВНИХ УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ І СКОРОЧЕНЬ	11
ВСТУП	12
РОЗДІЛ 1 АНАЛІЗ СУЧАСНОГО СТАНУ ДОСЛІДЖЕНЬ В ОБЛАСТІ АНАЛІЗУ ТОНАЛЬНОСТІ ТЕКСТОВОЇ ІНФОРМАЦІЇ ПРИ ВДОСКОНАЛЕННІ КОМП'ЮТЕРНИХ СИСТЕМ	16
1.1. Аналіз джерел одержання інформації про властивості комп'ютерних систем	16
1.1.1. Відкриті джерела текстової інформації та їх аналіз	16
1.1.2. Service Desk системи, як засіб вдосконалення комп'ютерних систем	17
1.2. Аналіз методів комп'ютерної лінгвістики при аналізі та оцінюванні тональності текстової інформації	22
1.3. Прикладні аспекти застосування комп'ютерної лінгвістики	25
1.4. Аналіз підходів до побудови лінгвістичних процесорів	33
1.5. Висновки до розділу	37
РОЗДІЛ 2 ПОБУДОВА МОДЕЛЕЙ ТА ОБГРУНТУВАННЯ МЕТОДІВ І МЕТРИК ОЦІНЮВАННЯ ТОНАЛЬНОСТІ ТЕКСТОВОЇ ІНФОРМАЦІЇ ПРИ ВДОСКОНАЛЕННІ КОМП'ЮТЕРНИХ СИСТЕМ	39
2.1. Обґрунтування та формалізація основних понять при використанні аспектного підходу виявлення та оцінювання тональності текстової інформації	39
2.2. Добування аспектів на основі частотних характеристик слів	43
2.3. Визначення аспектних термінів на основі фраз-порівнянь	46
2.3.1. Ітеративні методи для виявлення термінів-аспектів	46
2.3.2. Методи машинного навчання в задачах виявлення аспектів	48

- 2.3.3. Тематичні моделі для виявлення аспектних термінів50
- 2.4. Методи аналізу тональності текстової інформації при вдосконаленні комп'ютерних систем52
 - 2.4.1. Методи машинного навчання для аналізу тональності аспектів52
 - 2.4.2. Лінгвістично-інженерні методи оцінювання аспектів53
- 2.5. Методи і метрики оцінювання якості результатів класифікації55
 - 2.5.1. Метрика Accuracy55
 - 2.5.2. Метрика точності і повноти56
 - 2.5.3. Матриця контингентності58
 - 2.5.4. Застосування F-міри60
- 2.6. Висновки до розділу63

РОЗДІЛ 3 СИСТЕМА ВИЯВЛЕННЯ ТА АНАЛІЗУ ТОНАЛЬНОСТІ ТЕКСТОВОЇ ІНФОРМАЦІЇ ПРИ ВДОСКОНАЛЕННІ КОМП'ЮТЕРНИХ СИСТЕМ 63

- 3.1. Функціональні вимоги до системи виявлення та аналізу тональності текстової інформації64
- 3.2. Проектування схеми бази даних при визначенні тональності текстової інформації68
- 3.3. Проектування архітектури системи виявлення та оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем82
- 3.4. Висновки до розділу87

РОЗДІЛ 4 ОБГРУНТУВАННЯ ЕКОНОМІЧНОЇ ЕФЕКТИВНОСТІ88

- 4.1. Розрахунок норм часу на виконання науково-дослідної роботи88
- 4.2. Визначення витрат на оплату праці та відрахувань на соціальні заходи91
- 4.3. Розрахунок витрат на електроенергію93

- 4.4. Розрахунок витрат на матеріали94
- 4.5. Розрахунок суми амортизаційних відрахувань95
- 4.6. Обчислення накладних витрат96
- 4.7. Складання кошторису витрат та визначення собівартості науково-дослідних робіт96
- 4.8. Розрахунок ціни науково-дослідних робіт97
- 4.9. Визначення економічної ефективності і терміну окупності капітальних вкладень98

РОЗДІЛ 5 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ100

- 5.1. Охорона праці100
- 5.2. Запобігання забрудненню повітря, виробничих приміщень НХР, допустимі значення для основних речовин-забруднювачів; їх характеристика та засоби захисту103
- 5.3. Вплив іонізуючого випромінювання на користувачів комп'ютерних систем106

РОЗДІЛ 6 ЕКОЛОГІЯ109

- 6.1. Зниження енергоємності та енергозбереження109
- 6.2. Вимоги до мікроклімату, вмісту аеронів і шкідливих хімічних речовин у повітрі приміщень експлуатації моніторів і ПЕОМ. Шляхи дотримання цих вимог111

ВИСНОВКИ115

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ118

Додаток А Тексти наукових публікації122

ПЕРЕЛІК ОСНОВНИХ УМОВНИХ ПОЗНАЧЕНЬ,
СИМВОЛІВ І СКОРОЧЕНЬ

БД	База Даних
ЖЦ	Життєвий Цикл
КС	Комп'ютерна Система
ПП	Програмний Продукт
ПЗ	Програмне Забезпечення
CASE	Computer Aided Software Engineering
ER	Entity Relationships
UML	Unified Modeling Language

ВСТУП

Актуальність теми. Сучасний рівень розвитку інформаційних технологій зумовлює високу конкуренцію на ринку ІТ продуктів і послуг, а також змушує фірми-розробники автоматизувати практично усі бізнес-процеси. Наявність великої кількості даних та широке застосування методів інтелектуального аналізу текстової та графічної інформації, наявність апаратних і програмних засобів зберігання та опрацювання «великих даних» дає змогу майже в автоматичному режимі забезпечувати процеси комунікації між замовниками та розробниками, визначати переваги і недоліки наданих послуг чи проданих товарів, порівнювати конкурентоздатність продуктів, підвищити супровід комп'ютерних систем та їх вдосконалення.

Одним із найбільш перспективних напрямів, що сьогодні розвивається, є інтелектуальний аналіз відгуків користувачів про придбані товари, про використання тих чи інших апаратних засобів, програмного забезпечення і комп'ютерних систем в цілому. Розвиток цього напрямку зумовлений орієнтацією фірм-розробників на задоволеність користувачів і визначення конкурентних переваг чи недоліків в порівнянні з іншими продуктами. Визначення тональності відгуків користувачів про комп'ютерні системи та їх складові дає змогу визначити недоліки, які в подальшому будуть враховані при оновленні компонентів комп'ютерної системи. При цьому використовується методи семантичного та синтаксичного аналізу тексту, нейронні мережі, методи класифікації та кластеризації даних, що з певною точністю дозволяють визначити приналежність висловленої користувачем думки до одного з класів: позитивний, негативний чи нейтральний.

Дослідженню методів аналізу тональності і класифікації текстової інформації присвячено ряд наукових робіт як українських, так і закордонних вчених, зокрема Дж. Мак-Кіна, Г. Болла, Д. Холла, Г. Ланса, У. Уільямсона,

Н. Джардайна, Є.М. Бравермана, А.А. Дорофеюка, І.Б. Мучника, М.Г. Загоруйко та ін. Сьогодні ефективно застосовуються засоби аналізу тональності тексту, розроблені фірмами Google, Yahoo та Yandex.

На сьогодні розроблено багато методів для вирішення задач щодо визначення тональності тексту. Однак, ефективність їх застосування залежить від природи вхідних даних, області та способів використання. Обґрунтування і розробка методу визначення тональності текстової інформації при вдосконаленні комп'ютерних систем є однією з нових та актуальних задач і вимагає додаткового дослідження існуючих методів і засобів.

Метою роботи є дослідження методів і засобів визначення тональності текстової інформації для вдосконалення комп'ютерних систем.

Для досягнення вказаної мети в роботі поставлено наступні **задачі**:

- аналіз наукових праць і практик реалізації методів і засобів виявлення та аналізу тональності текстової інформації;
- аналіз особливостей опрацювання відгуків користувачів при вдосконаленні комп'ютерних систем;
- обґрунтування інтелектуальних методів аналізу та визначення тональності у відгуках користувачів про комп'ютерні системи;
- розробка інтелектуального методу виявлення тональності відгуків користувачів;
- розробка програмного засобу виявлення та оцінювання тональності текстової інформації при супроводі комп'ютерних систем.

Об'єкт дослідження: процес виявлення тональності текстової інформації.

Предмет дослідження: методи і засоби виявлення та аналізу тональності текстової інформації.

Методи дослідження. Для вирішення поставлених задач використано наступні методи: аналіз та узагальнення – при проведенні аналізу існуючих методів і засобів виявлення тональності текстової інформації; теорії

імовірності та математичної статистики, машинного навчання – при розробці методу виявлення та аналізу тональності відгуків користувачів комп'ютерних систем; проектування та програмування – при розробці програмного засобу збору, аналізу та виявлення тональності текстової інформації при супроводі комп'ютерних систем; експеримент і тестування – для апробації розробленого методу та програмного засобу.

Наукова новизна одержаних результатів:

– уперше запропоновано архітектурне рішення та спроектовано базу даних програмного засобу, що оперує поняттями лінгвістично-інженерного підходу та машинного навчання і дає змогу автоматизувати процес оцінювання тональності текстової інформації та приймати рішення щодо вдосконалення властивостей комп'ютерних систем;

– набули подальшого розвитку методи і засоби аналізу текстової інформації при вдосконаленні комп'ютерних систем, в основі якого лежить поєднання лінгвістично-інженерного підходу та машинного навчання у контексті застосування шаблонів синтаксичних конструкцій, методів машинного навчання з учителем, що дає змогу виявляти та оцінювати аспекти комп'ютерних систем із заданою точністю і повнотою.

Практичне значення одержаних результатів. Запропонований підхід виявлення та оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем впроваджено у вигляді програмного інструменту, який дає змогу виявляти аспекти у відгуках користувачів, групувати їх та кількісно виражати ступінь задоволеності користувачів щодо реалізованих у комп'ютерній системі властивостей.

Публікації. Результати дослідження апробовано на VIII міжнародній науково - технічній конференції молодих учених і студентів «Актуальні задачі сучасних технологій» (27-28 листопада 2019 р.) Тернопільського національного технічного університету імені Івана Пулюя та на VII науково-технічній конференції Тернопільського національного технічного

університету імені Івана Пулюя «Інформаційні моделі, системи та технології» (11-12 грудня 2018 року) у вигляді тез конференцій:

1. Гаван Р.Р., Яцишин В.В. Підходи до вдосконалення комп'ютерних систем на основі аналізу тональності відгуків користувачів. Матеріали VIII Міжнародної науково-технічної конференції молодих учених та студентів «Актуальні задачі сучасних технологій» (27 - 28 листопада 2019 р.). Тернопіль. 2019. с. 16

2. Яцишин В.В., Гаван Р.Р. Особливості аналізу тональності тексту на основі аспектів при вдосконаленні комп'ютерних систем. Матеріали VII науково-технічної конференції Тернопільського національного технічного університету імені Івана Пулюя «Інформаційні моделі, системи та технології» (11-12 грудня 2019 року). Тернопіль, ТНТУ. 2019. с. 114

Структура роботи. Робота складається з розрахунково-пояснювальної записки та графічної частини. Розрахунково-пояснювальна записка складається із вступу, 6 розділів, висновків, переліку посилань та додатків. Обсяг роботи: розрахунково-пояснювальна записка – 125 арк. формату А4, графічна частина – 12 аркушів формату А1.

РОЗДІЛ 1
АНАЛІЗ СУЧАСНОГО
СТАНУ ДОСЛІДЖЕНЬ
В ОБЛАСТІ АНАЛІЗУ
ТОНАЛЬНОСТІ
ТЕКСТОВОЇ
ІНФОРМАЦІЇ ПРИ
ВДОСКОНАЛЕННІ
КОМП'ЮТЕРНИХ
СИСТЕМ

1.1. Аналіз джерел одержання
інформації про властивості
комп'ютерних систем

1.1.1. Відкриті джерела текстової інформації та їх аналіз

Відгуки користувачів про придбаний товар, одержані послуги, поведінку чи компоненти комп'ютерної системи є визначальними для будь-якої сфери діяльності людини. У реальному світі підприємства та організації завжди хочуть знати думку споживачів чи громадськості щодо своїх товарів та послуг. Конкретні споживачі також хочуть дізнатися думку інших користувачів товару перед його придбанням, а також думки інших щодо політичних кандидатів, перш ніж приймати рішення про голосування на виборах. Коли організації чи бізнес-компанії хочуть одержати відгуки, то вони проводять опитування у фокус-групи.

Аналіз та оцінювання думок громадськості та споживачів вже давно є величезним бізнесом для маркетингу, зв'язків з громадськістю та політичних кампаній. Із вибуховим зростанням соціальних медіа (наприклад, блоги, мікро-блоги, Twitter, коментарі та публікації на сайтах соціальних мереж) у мережі люди та організації все частіше використовують вміст у цих ЗМІ для прийняття рішень.

Із розвитком соціальних медіа організаціям не потрібно проводити опитування, оскільки існує велика кількість такої інформації, яка є у

відкритому доступі. Однак пошук і моніторинг ресурсів у мережі Інтернет та аналіз одержаної інформації, залишається трудозатратним завданням. Кожен ресурс, зазвичай, містить величезний обсяг відгуків, які не завжди легко розшифрувати у довгих текстах та публікаціях на форумах.

Таким чином, для вирішення задач виявлення та аналізу думок і відгуків користувачів необхідні автоматизовані системи аналізу настроїв. Останніми роками спостерігаються тенденції міграції бізнесу у соціальні мережі, а також застосування їх як інструментів впливу на виборі кампанії.

Окрім відкритих даних, великі компанії володіють власними внутрішніми базами даних, які зазвичай формуються для підтримки кінцевих користувачів. До таких даних належать відгуки клієнтів, зібрані з електронних листів та кол-центрів, або результати опитувань, проведених організаціями. Завдяки цим додаткам спостерігається зростання користувачів комп'ютерних систем – лише в США було створено 40-60 стартап-компаній по аналізу тональності текстової інформації. Багато великих корпорацій також створили власні внутрішні можливості, наприклад, Microsoft, Google, HewlettPackard, SAP та SAS. Ці практичні додатки та промислові інтереси мотивують дослідження аналізу настроїв користувачів.

1.1.2. Service Desk системи, як засіб вдосконалення комп'ютерних систем

Впровадження служби Service Desk зазвичай здійснюється у вигляді проекту, який передбачає попередній консалтинг. Якісна реалізація такого проекту можлива тільки при достатній кваліфікації фахівців, що беруть участь у впровадженні. Більшість розробників комп'ютерних систем, орієнтованих на реалізацію служб на зразок Service Desk, не рекомендують впроваджувати рішення за відсутності у виконавців сертифікатів, які гарантують їх кваліфікацію.

Налагодження ефективної технічної підтримки є пріоритетним завданням конкурентної стратегії компанії. При цьому для планування, побудови і подальшого управління сервісом необхідна можливість його описати і виміряти за допомогою кількісних показників.

Існують різні моделі та відповідні метрики, призначені для опису процесів підтримки, що появилися в різних областях (робота Call-центру, IT-служби підприємства та ін.) Це накладає певні обмеження на використання кожної з цих моделей в якості універсального засобу опису сервісу.

Визначення основних концептуальних умов функціонування служби реєстрації та обробки проблемних ситуацій є хорошою основою для підвищення продуктивності системи після її реалізації. Цей процес також дозволяє уникнути помилок ще на етапі початкового проектування архітектури майбутньої системи.

Представлення технічної підтримки в якості окремого бізнес-процесу є досить спрощеним, тому що дане бізнес-явище включає в себе безліч елементів діяльності організації.

На рис. 1.1 представлено типовий ланцюжок створення цінності в організації. До технічної підтримки відносяться компоненти, розташовані на всіх рівнях, а саме:

- основні та допоміжні бізнес-процеси;
- основні засоби для забезпечення технічної підтримки користувачів;
- спеціалізовані інформаційні системи та обладнання (Service Desk, ПЗ Contact-центру та ін);
- персонал;
- CRM-система, що узгоджує інформацію про надання сервісу в загальне сховище даних із взаємин із споживачами, а також з іншими ІС підприємства (ERP і ін);

- бізнес-процеси управління технічною підтримкою користувачів, методики її планування та оцінки ефективності;
- відповідний розділ стратегії компанії, що описує політику в області технічної підтримки.

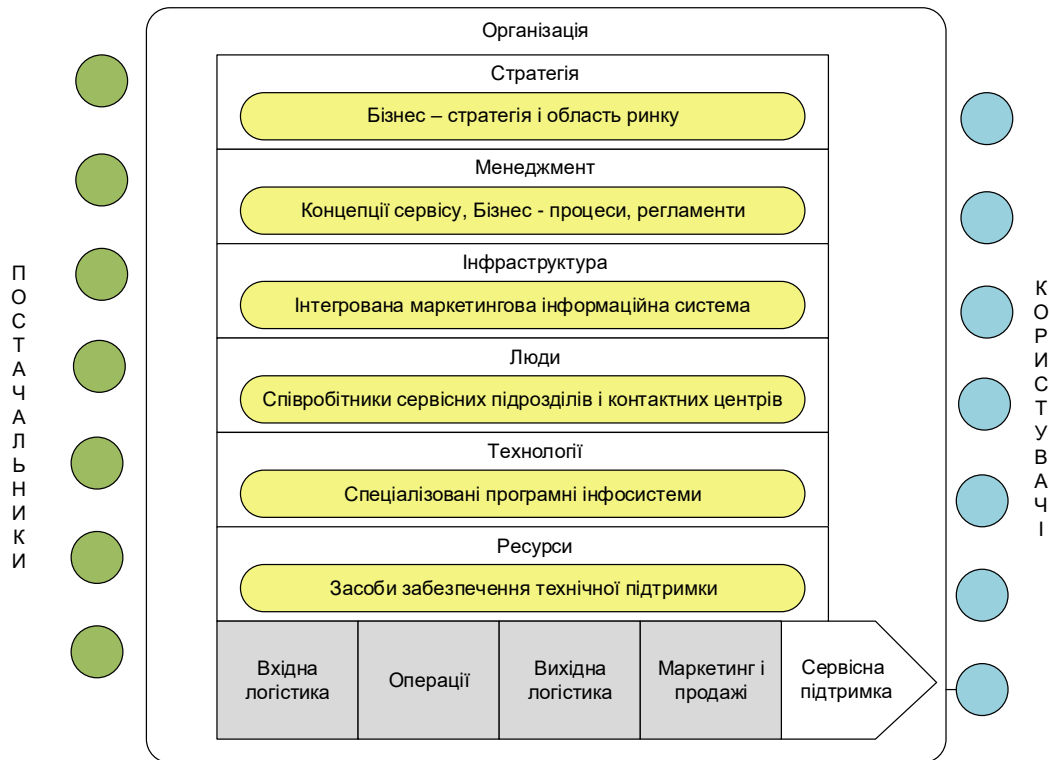


Рис. 1.1. Схема забезпечення функціонального циклу організації

Таким чином, технічну підтримку клієнтів можна визначити як частину діяльності організації, спрямовану на надання допомоги споживачам в процесі використання ними придбаних продуктів або послуг, у відповідності з місією і стратегічними цілями даної організації.

Розглянуті нижче моделі (рис. 1.2) виходять за межі даного визначення, що не включають в себе процеси продажів (в т.ч. продажів додаткових продуктів і послуг - Up-Cross-Selling), маркетингу, передпродажної підготовки (Presale). В рамках технічної підтримки взаємодія відбувається з уже існуючим клієнтом і виключно з питань, пов'язаних з використанням вже придбаного ним обладнання та послуг. При цьому технічна і сервісна підтримка може надаватися як на безплатній

(підтримка абонентів телекомунікаційних компаній), так і на платній основі (сервісне обслуговування автомобілів).

Крім того, слід поглянути на технічну підтримку не тільки як на важливу частину самої організації, а й як на важливу складову цінової пропозиції її продуктів і послуг.

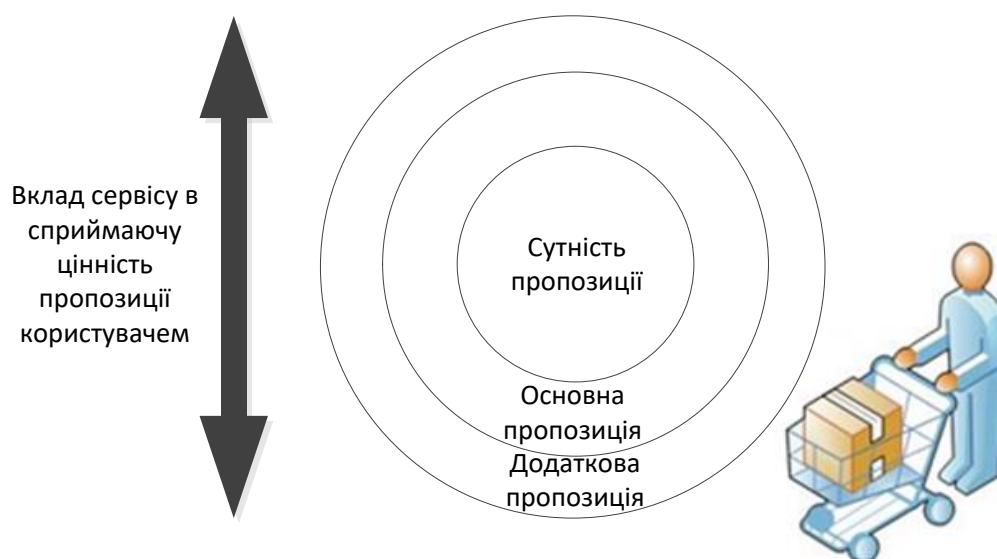


Рис. 1.2. Маркетинговий цикл продукту

Розробляючи новий продукт, менеджер і маркетолог орієнтуються на задовольняючі з його допомогою потреби покупця (сутність пропозиції), при цьому він додає йому певний набір зовнішніх якостей – форма, колір, упаковка (основна пропозиція) і набір додаткових властивостей – спосіб покупки, після продажне обслуговування (додаткова пропозиція). Сукупність всіх цих компонентів у зв'язку з ціновою політикою утворює загальну цінову пропозицію, яка, очевидно, має бути спочатку краща, ніж у конкурентів.

Параметри технічної підтримки можуть виступати в даному випадку в якості компонентів на всіх трьох рівнях цінової пропозиції, надаючи йому додаткові конкурентні переваги, наприклад:

- технічна і сервісна підтримка сама по собі може бути продуктом – послуги аутсорсингового контакт-центру по обслуговуванню клієнтів зовнішніх компаній;

- технічна і сервісна підтримка може бути способом пропозиції основного продукту – покупці деяких програмних продуктів отримують права на їх використання безкоштовно, але укладають договори на консультації, вирішення проблем в процесі їх використання з компанією-розробником;

- технічна і сервісна підтримка є доповненням до основної пропозиції (найбільш поширений випадок) – безкоштовне технічне обслуговування автомобіля, веб-портал самообслуговування власників банківських карт.

Всі ці приклади показують, що добре організовану і високотехнологічну підтримку клієнтів можна використовувати для надання продуктам і послугам компанії додаткових конкурентних переваг в самих різних аспектах. Природно, для цього технічна підтримка повинна бути прозора з точки зору організації, визначена і керована.

Крім того, використання технічної підтримки всередині інфраструктури підприємства допомагає підвищити зв'язок між підрозділами, які створюють основний продукт і обслуговуючим персоналом. Тим самим існування єдиної системи збору інформації про інциденти, диференціальний розподіл завдань між обслуговуючими підрозділами, накопиченні бази знань за рішенням різних проблем підвищують швидкість обробки таких повідомлень, зменшує час реакції відповідних служб, тим самим допомагаючи заощаджувати фінансові і часові ресурси. Крім цього, такі системи є джерелом для аналізу властивостей комп'ютерних систем і шляхів їх вдосконалення.

1.2. Аналіз методів комп'ютерної лінгвістики при аналізі та оцінюванні тональності текстової інформації

Розвиток мережі Інтернет і стрімке зростання кількості текстової інформації значно прискорює і стимулює розвиток науково-прикладної області, яка існує вже протягом десятків років – автоматична обробка текстів (Natural Language Processing) та комп'ютерна лінгвістика (Computational Linguistics). У цій галузі запропоновано багато перспективних ідей щодо автоматичного опрацювання текстів природньою мовою, які були впроваджені у багатьох прикладних системах, зокрема, комерційних. Сфера застосування комп'ютерної лінгвістики постійно розширюється, з'являються нові задачі, які успішно розв'язуються, в тому числі з використанням результатів з інших наукових областей. Інформацію про наукові досягнення у даній галузі агреговані на сайті ACL (Association of Computational Linguistics) [1].

Комп'ютерна лінгвістика представляє собою міждисциплінарну область, яка виникла на межі таких наук, як лінгвістика, математика, інформатика, штучний інтелект. Для комп'ютерної лінгвістики характерне використання та адаптація розроблених у даних науках методів і засобів.

Перші праці комп'ютерної лінгвістики написані американським лінгвістом Н. Хомським щодо формалізації структури природної мови. Окрім цього, вчений провів експеримент машинного перекладу, який виконав за допомогою досягнень математиків та програмістів, а також до розроблених в області штучного інтелекту першими програмами розуміння природної мови.

Об'єктом дослідження у комп'ютерній лінгвістиці є процеси опрацювання природної мови, тому її розвиток неможливий без базових знань загальної лінгвістики [2]. Лінгвістика вивчає загальні закони

природної мови – її структуру та функціонування, і включає області, які наведені у табл. 1.1.

Області лінгвістики

Область лінгвістики	Опис
Фонетика	вивчає звуки мови та правила їх з'єднання при формуванні мовних одиниць
Морфологія	досліджує внутрішню структуру і зовнішню форму слів у мові, включаючи частини мови та їх категорії
Синтаксис	вивчає структуру речень, правил їх побудови і порядку вживання слів у реченні, а також загальні властивості речень, як одиниць мови
Семантика	досліджує зміст слів, речень та інших одиниць мови
Лексика	описує лексикон конкретної природної мови – її окремі слова, їх граматичні та семантичні властивості, а також методи створення словників

Найбільш тісно комп'ютерна лінгвістика пов'язана з областю штучного інтелекту, в рамках якої розробляються програмні моделі окремих інтелектуальних функцій. Незважаючи на очевидний перетин комп'ютерної лінгвістики та штучного інтелекту (володіння мовою відноситься до інтелектуальних функцій), штучний інтелект не поглинає всю комп'ютерну лінгвістику, оскільки вона має свій теоретичний базис і методологію. Спільним для цих наук є комп'ютерне моделювання як основний спосіб та кінцева ціль дослідження, евристичний характер багатьох наявних методів.

У спрощеному вигляді, завдання комп'ютерної лінгвістики можуть бути сформульовані як розробка методів і засобів побудови лінгвістичних процесорів для різних прикладних задач щодо автоматичного опрацювання текстів природною мовою. Розробка лінгвістичного процесора для деякої прикладної задачі передбачає формальний опис лінгвістичних властивостей при опрацюванні текстів, який може розглядатись як модель тексту (або модель мови).

1.3. Прикладні аспекти застосування комп'ютерної лінгвістики

Область застосування комп'ютерної лінгвістики постійно розширюється, тому наведемо найбільш широко розв'язувані прикладні задачі, які використовуються її інструменти.

Машинний переклад – найбільш раннє застосування комп'ютерної лінгвістики, разом з яким виникала і розвивалась сама область. Перші програми перекладу природної мови були створені у середині минулого століття, в основі яких лежав принцип простого перекладу слів у тексті. Однак доволі швидко, стало зрозуміло, що машинний переклад вимагає значно більш повної лінгвістичної моделі. Така модель була розроблена у системі ЕТАП [3], а також у кількох інших системах, які виконують переклад наукових текстів.

На сьогоднішній день існує цілий спектр комп'ютерних систем машинного перекладу, від великих інтернаціональних дослідницьких проектів до комерційних автоматичних перекладачів. Найбільшу зацікавленість становлять проекти мультимовного перекладу з використанням проміжних проміжної мови, за допомогою якої кодується зміст фраз, які підлягають перекладу.

Сучасний тренд у машинному перекладі – статистична трансляція, що опирається на статистику пар слів і словосполучень, які перекладаються.

Незважаючи на тривалість досліджень і розвиток методів перекладу природної мови, якість перекладу все ж залишається на не надто високому рівні. Однак вагомий прогрес в області відбувається завдяки використанню штучних нейронних мереж. Приклад організації і застосування машинного перекладу наведено на рис. 1.3.

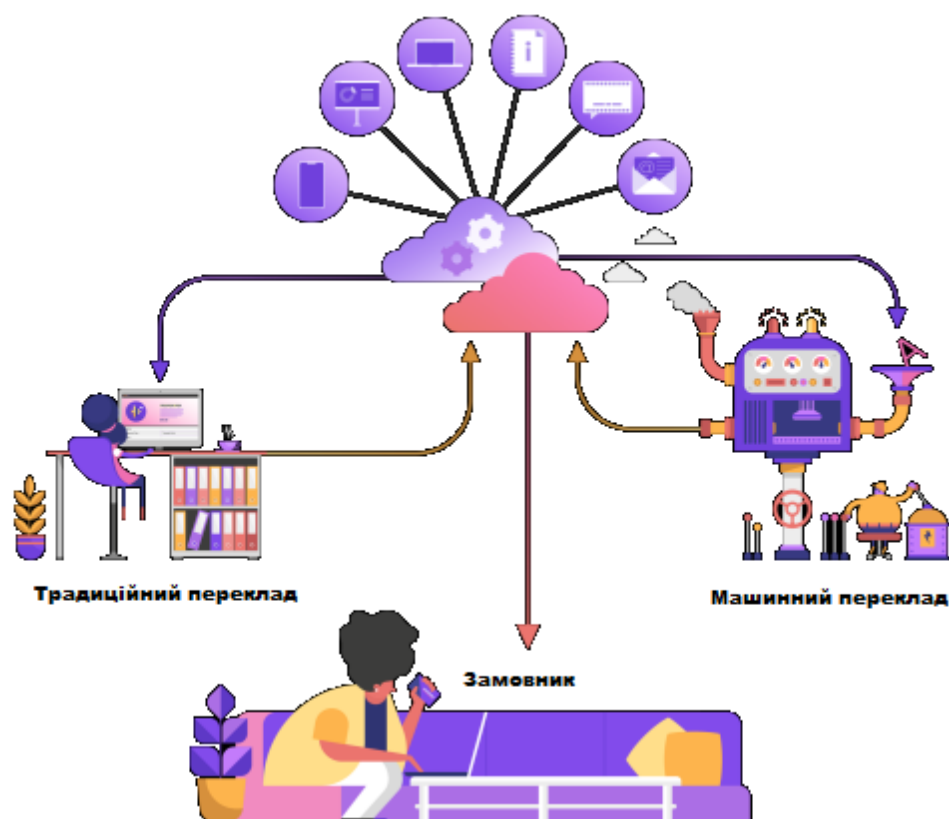


Рис. 1.3. Схема організації машинного перекладу

Іншою сферою застосування комп'ютерної лінгвістики є інформаційний пошук (Information Retrieval) та пов'язані з ним задачі індексування, реферування, класифікації і рубрикації документів.

Повнотекстовий пошук документів у великих базах текстових документів (Big Data) передбачає їхнє індексування та вимагає попереднього опрацювання і створення індексних груп. Відомо кілька ефективних моделей інформаційного пошуку, зокрема, найбільш використовуваною на даний час є векторна модель. При застосуванні даної

моделі інформаційний запит представляється у вигляді набору слів, а відповідні релевантні документи визначається на основі застосування метрик подібності запиту і вектора слів у документі. Сучасні системи пошуку в мережі Інтернет реалізують цю модель, і виконують індексацію текстів за словами, які у них зустрічаються, а при видачі результатів пошуку використовують специфічні процедури та алгоритми ранжування документів. Актуальною задачею для цієї області є мультимовний пошук у документах.

Реферування тексту (Summarization) є однією із прикладних областей комп'ютерної лінгвістики і вирішує задачу, суть якої полягає у скороченні довжини тексту та одержанні його короткого змісту – реферату (анотації). На рис. 1.4 візуалізовано процес реферування.



Рис. 1.4. Реферування

Це дає змогу підвищити продуктивність пошуку тексту у колекціях документів. Реферат може формуватися для також для кількох подібних за темою документів (наприклад кластер новин). Основним методом автоматичного реферування до цих пір є відбір найбільш важливих та вагомих речень у тексті, що реферується, на основі статистики слів і словосполучень, а також структурних і лінгвістичних особливостей тексту.

Близька до реферування задача – анотація текстових документів. У найпростішому вигляді анотація представляє собою перелік основних

ключових тем тексту, для добування яких використовуються статистичні і лінгвістичні критерії.

При опрацюванні великих колекцій документів актуальними також є задачі класифікації і кластеризації текстів. Класифікація представляє собою процес визначення приналежності кожного документу до визначеного класу із наперед заданими параметрами, а кластеризація – процес розбиття множини документів на кластери, тобто формування підмножин тематично близьких документів.

Для розв'язку таких задач застосовуються методи машинного навчання. У зв'язку з цим задачі класифікації і кластеризації текстової інформації належать до напрямку Text Mining, що є частиною Data Mining – область інтелектуального аналізу даних.

Прикладна область задач класифікації дає змогу вирішувати розпізнавання спаму, класифікацію SMS-повідомлень та ін.

Дуже близькою до класифікації є задача рубрикації тексту (Text Classification) – визначення приналежності тексту до однієї з наперед відомих тематичних груп, які зазвичай, утворюють ієрархічне дерево тем.

Відносно новою задачею із застосуванням класифікації є задача інформаційного пошуку – формування відповідей на запитання (Question Answering) [4]. Розв'язок такої задачі одержується шляхом визначення типу питання і пошуком текстів, які можуть потенційно містити відповідь на поставлене запитання і після цього добування відповіді із знайдених текстів.

Актуальна на сьогодні прикладна задачею, що відноситься до напрямку Text Mining, стосується ідентифікації та добування інформації з тексту (Information Extraction) [5]. Такий тип задач притаманний економічній та виробничій аналітиці. При розв'язку даної задачі виконується добування у тексті, написаному на природній мові, певних визначених об'єктів – іменованих сутностей (імен осіб, географічних назв, назв компаній і т.п.) та їх зв'язків з деякими подіями. Як правило, це реалізується на основі часткового синтаксичного аналізу тексту, що дає змогу виконувати

опрацювання великих масивів текстів, зокрема, потоків новин від інформаційних агенств. Визначені дані певним чином структуруються і/або візуалізуються.

До напряму досліджень Text Mining відносять ще дві задачі – виявлення думок (Opinion Mining) та аналіз тональності текстів (Sentiment Analysis), які сьогодні є дуже актуальними задачами і досліджуються багатьма вченими. При вирішенні першої задачі щодо думок користувачів відбувається їх пошук у тематичних блогах, форумах, інтернет-магазинах про товар або інші об'єкти та проводиться їх аналіз.

Задача виявлення тональності текстової інформації близька до задачі контент-аналізу текстів масової комунікації при якій оцінюється загальна тональність висловлювань і тексту в цілому.

Ще однією задачею, яка виникала близько 50 років тому, і активно розв'язується із появою та розвитком мережі Інтернет полягає у підтримці діалогу на природній мові. До появи Інтернету ця задача розв'язувалась в рамках деякої інформаційної системи, зокрема, при опрацюванні запитів природною мовою до спеціалізованої бази даних. При цьому слова у запитах і базі даних були достатньою обмеженою множиною в лексичному і граматичному плані, що давало змогу використовувати спрощені методи аналізу запитів, а відповіді будувати за шаблонами. На даний час широкої популярності набувають чат-боти, які здатні підтримувати розмову з людиною на деяку тематику. Чат-боти є нащадками відомої системи ELIZA і проходять відомий тест Тьюринга.

Інша задача, що розв'язується в рамках комп'ютерної лінгвістики, полягає в автоматизації підготовки і редагування текстів природною мовою. Одними з перших досягнень у даній сфері були створенні програми орфографічної перевірки тексту, зокрема автокоректори і «спеллери». Перевірка орфографії уже давно реалізована у комерційних системах, однак для неї характерні частотні синтаксичні помилки. В автокоректорах поки що не реалізовано розпізнавання більш складних помилок, зокрема,

неправильне вживання присудків і лексичні помилки, що виникають у результаті набору тексту або не вірного вживання подібних слів. Сучасні дослідження комп'ютерної лінгвістики спрямовані на розвиток методів автоматизованого виявлення та виправлення помилок, що базуються на статистиці вживання слів і словосполучень.

Іншою прикладною задачею у сфері комп'ютерної лінгвістики є навчання природної мови. У рамках цього напрямку створюються комп'ютерні системи, що підтримують вивчення окремих аспектів (морфології, лексики, синтаксису) мови – англійської, російської та ін (Duolingo, LingvoLeo та ін.). Розробляються також багатофункціональні комп'ютерні словники, які не мають текстових аналогів і орієнтовані на широке коло користувачів, наприклад, словник сполучення слів для російської мови КроссЛексика, додатково формує синоніми, антоніми та інші змістові зв'язки для слів.

Наступний прикладний напрям, який варто проаналізувати – це автоматична генерація текстів природною мовою [2]. В принципі, це завдання можна вважати підзадачею вже розглянутої вище задачі машинного перекладу, проте в рамках напрямку є ряд специфічних завдань. Такою задачею є багатомовна генерація, тобто автоматична побудова відразу на декількох мовах спеціальних документів – патентних формул, інструкцій з експлуатації технічних засобів або програмних систем, виходячи з їх формальної специфікації. Прикладом такої системи є Crowdin.

Одним з напрямків досліджень, який активно розвивається, є розпізнавання і синтез голосової мови. Помилки, які неминуче виникають при розпізнаванні, виправляються автоматичними методами на основі словників і морфологічних моделей із застосуванням методів машинного навчання. Складність моделювання у комп'ютерній лінгвістиці пов'язана з тим, що природна мова представляє собою велику відкриту багаторівневу систему знаків для обміну інформацією в процесі практичної діяльності людини, і постійно змінюється у зв'язку з цією діяльністю [6,7].

Текст на природній мові складається з окремих елементів (знаків), і при цьому, можливі декілька способів декомпозиції тексту на структурні елементи, що відносяться до різних рівнів.

Загально визнаним є існування наступних рівнів [9]:

- рівень речень (висловлювань) - синтаксичний рівень;
- рівень слів (словоформ - слів в певній граматичній формі, наприклад, ручка, дружбою) - морфологічний рівень;
- рівень фонем (окремих звуків, за допомогою яких формуються і розрізняються слова) - фонетичний рівень.

Фонетичний рівень використовується при усному мовленні, а для письмових текстів в мовах з алфавітним способом запису (зокрема, в європейських мовах) він відповідає рівню символів (звуки приблизно відповідають буквам алфавіту).

Рівні, по суті, є підсистеми загальної системи природної мови (взаємопов'язані, але в достатній мірі автономні), і вони самі можуть містити підсистеми. Так, морфологічний рівень включає також підрівень морфем. Морфема – це мінімальна значуща частина слова (корінь, префікс, суфікс, закінчення, постфікси).

Питання кількості рівнів у лінгвістиці досі залишається відкритим. Як окремий, може бути виділений лексичний рівень – рівень лексем.

Лексема – це слово, як сукупність всіх його конкретних граматичних форм (наприклад, лексему лист утворюють форми лист, листа, листу, листом). Точніше, лексема – семантичний інваріант всіх словоформ.

У тексті зустрічаються словоформи (лексеми в певній формі), а в словнику природної мови – лексеми, у словнику записується канонічна словоформа лексеми, яку часто називають лемою (наприклад, для іменників це форма називного відмінка однини: лист).

У рамках синтаксичного рівня може бути виділений підрівень словосполучень – синтаксично пов'язаних груп слів (бачив ліс, синя куля),

і надрівень складного синтаксичного цілого, якому приблизно відповідає абзац тексту.

Складне синтаксичне ціле – це послідовність пропозицій (висловлювань), об'єднаних змістом і лексико-граматичними засобами [10]. До таких засобів відносяться в першу чергу лексичні повтори – посилання на попередні слова тексту, які реалізуються за допомогою займенників і займенникових слів (вони, цей, там же і т. д.).

Ієрархія рівнів проявляється в тому, що одиниці більш високого рівня розкладаються на одиниці нижчого рівня (наприклад, словоформи на морфеми). Вищий рівень великою мірою зумовлює організацію нижчого рівня. Наприклад, синтаксична структура речення у більшості випадків визначає, які повинні бути обрані словоформи.

Можна також говорити ще про рівні дискурсу [11], під яким розуміється зв'язний текст його комунікативної спрямованості. Під дискурсом розуміється послідовність взаємопов'язаних між собою речень тексту, що володіє певною змістовою цілісністю, за рахунок чого він виконує певну прикладну задачу. У багатьох типах зв'язних текстів проявляється традиційна схематична (дискурсивна) структура, яка організовує їх загальний зміст, наприклад, певну структуру мають опис складних технічних систем, патентні формули, наукові статті, ділові листи та ін.

Особливим є питання про рівень семантики. В принципі, сенс є всюди, де є знакові одиниці мови (морфеми, слова, речення). Підтвердженням самостійності рівня семантики вважається те, що людина зазвичай запам'ятовує сенс висловлювання, а не його конкретну мовну форму. До сих пір не зрозумілою є організація цього рівня. Передбачається, що існує універсальний набір елементарних семантичних одиниць, приблизно 2 тисячі, за допомогою яких можна висловити зміст будь-якого висловлювання.

Крім багаторівневості системи природної мови, складність її моделювання пов'язана з постійним змінами у ній. Зміни стосуються не тільки словникового запасу мови (нові слова і новий зміст старих), але також синтаксису, морфології і фонетики. Як наслідок, принципово неможливо один раз розробити формальну модель конкретної природної мови і побудувати відповідний лінгвістичний процесор. Потрібне постійне поповнення знань про мову на всіх рівнях і корекція існуючих моделей.

Одним із наслідків тривалого історичного розвитку природних мов є нестандартна сполучність (синтактика) одиниць на кожному рівні мови. На відміну від штучних формальних мов (мов логіки, мов програмування), в яких сполучність знаків диктується їх семантикою та може бути зафіксована синтаксично (граматично), в природних мовах з'єднання слів у реченнях лише частково може бути описана законами граматики.

У будь-якій мові досить багато граматично правильних з'єднань реально не вживається, наприклад, в українській мові є поєднання міцний чай, але не важкий чай (як в англійській мові strong tea).

Однією з найбільших проблем при опрацюванні текстів на природній мові є неоднозначність та багатозначність її конструктивних елементів, що проявляється на усіх його рівнях і виражається у явищах полісемії, омонімії, синонімії.

1.4. Аналіз підходів до побудови лінгвістичних процесорів

На сьогодні для створення модулів лінгвістичних процесорів застосовується два основних підходи: заснований на правилах (rule based), або інженерний, і заснований на машинному навчанні (machine learning).

Історично першим є підхід, що використовує правила. Суть підходу полягає в описі необхідної лінгвістичної інформації у вигляді формальних правил. У ранніх системах правила були вбудовані в програмний код, зараз

же для запису правил використовується або вже готову формальну мову, або мову спеціально створюють для системи, що розробляється.

Правила створюються лінгвістами або фахівцями з проблемної області текстів. В рамках підходу, заснованого на машинному навчанні, джерелом лінгвістичної інформації виступають не правила, а відібрані тексти проблемної області. Серед методів, що застосовуються в рамках підходу, виділяють методи навчання з учителем (supervised), методи навчання без вчителя (unsupervised), методи часткового навчання з учителем (bootstrapping).

Найчастіше застосовується навчання з учителем, при якому відбувається побудова математичної і програмної моделі – машинного класифікатора, який вміє розпізнавати різні класи одиниць тексту (слів, словосполучень та інших конструкцій) або самих текстів. Побудова класифікатора відбувається на спеціально розміченому текстовому корпусі (навчальній вибірці), в якому, одиницям, що розпізнаються, приписані мітки, які кодують важливі ознаки елементів/текстів. Навчання представляє собою виявлення загальних закономірностей, властивих текстам на природній мові, де за основу беруться дані навчальної вибірки.

Обидва розглянутих підходи мають свої переваги і недоліки. Створення правил вимагає значних трудозатрат і вимагає досить кваліфікованої праці, як правило, лінгвіста. Дуже часто навіть лінгвіст не може передбачити заздалегідь всі окремі випадки, які треба відобразити у правилах.

У той же час правила зазвичай є декларативними і легко зрозумілі, тому їх просто підтримувати: модифікувати і розширювати, тим самим налагоджуючи функціонування процесора.

Машинне навчання не вимагає ручної праці по складанню правил і скорочує час розробки систем, проте необхідні знання для вибору відповідних методів навчання. Крім того, результуючі моделі (класифікатори) непрозорі для розуміння, оскільки не мають явної

лінгвістичної інтерпретації (рис. 1.5 та рис. 1.6). Також машинне навчання передбачає наявність відповідного розміченого корпусу текстів, що не завжди можливо. Створення такого корпусу в будь-якому випадку вимагає значних обсягів ручної праці.



Рис. 1.5. Методи машинного навчання із вчителем



Рис. 1.6 Методи навчання без вчителя

Порівнюючи застосування цих підходів, можна помітити, що раніше більш часто застосовувався підхід на правилах, оскільки було мало розмічених текстових корпусів. З появою різних розмічених даних все

частіше вдаються до машинного навчання, як швидкого способу отримання потрібного додатка комп'ютерної лінгвістики.

Сучасна тенденція – модульні, багатокomпонентні системи автоматичної обробки текстів (multi-component, pipelined systems), причому різні модулі можуть бути створені в рамках різних підходів, наприклад, модуль графематичного аналізу – на основі машинного навчання, а морфологічного – на основі правил.

Машинне навчання досить часто застосовується для опрацювання колекцій текстових документів, з використанням моделі ознак тексту, при цьому ознаки визначені для кожного документа окремо. Ознаками можуть виступати різні інформаційні характеристики тексту: як лінгвістичні, так статистичні та структурні: наприклад, частота певних слів (або їх категорій) в документі, частота використання спецзнаків, співвідношення частин мови, наявність певних синтаксичних конструкцій або розділів тексту, дата створення та ін.

Різновидами моделі ознак є модель BOW (bag of words – мішок слів), в якій текст характеризується набором своїх значущих слів, а також векторна модель тексту, в якій вказаний набір впорядкований (рис. 1.7). Векторна модель застосовується, наприклад, при інформаційному пошуку. В якості ознак частіше беруться не слова, а більш складні характеристики, такі як показник TF-IDF [12] для слів (рис. 1.8).

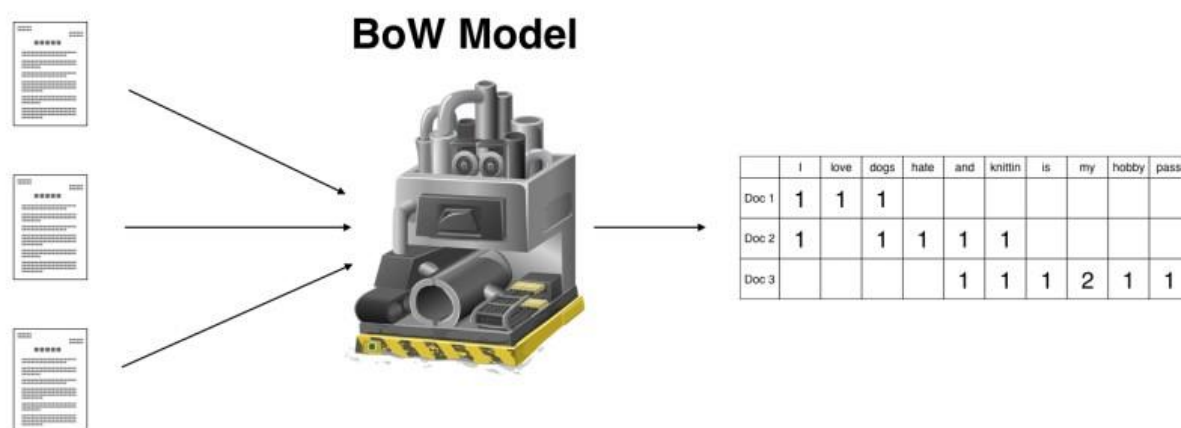


Рис. 1.7. Алгоритм Bag-of-words

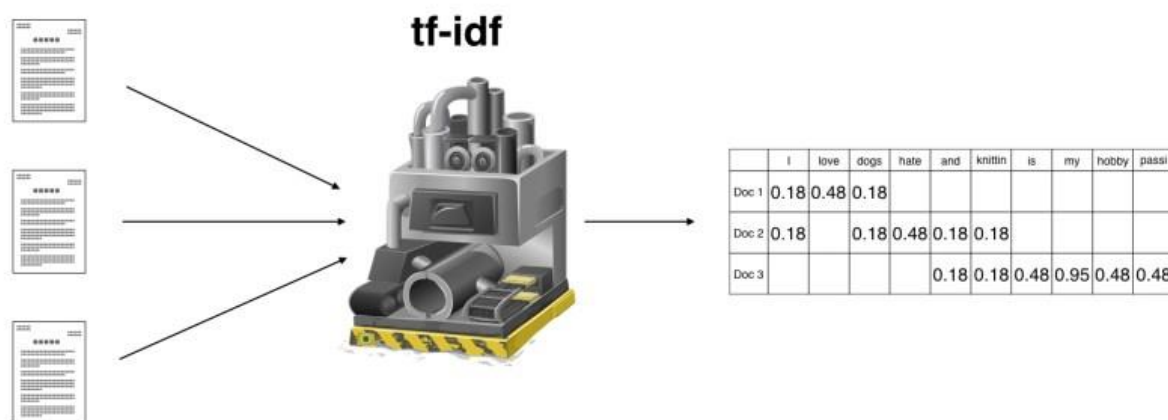


Рис. 1.8. Алгоритм TF-IDF

Окремо виділяють статистична мовну модель (Language Model), що характеризує мову в цілому, а не окремий текст [13, 14].

Класична мовна модель будується за масивом текстів конкретної природної мови (наприклад, англійської) шляхом підрахунку частот -грам слів (тобто тих слів, що стоять поруч слів). Найчастіше розглядаються біграми і триграми.

Модель покликана давати відповідь на питання, наскільки ймовірною є поява заданого слова, якщо безпосередньо перед ним зустрічалися певні слова. Ймовірності розраховуються на основі зібраної статистики. Така модель застосовується, наприклад, для вирішення лексичної неоднозначності.

Різновиди моделі: N-грами частин мови слів тексту або N-грами букв тексту (можливі й інші моделі) застосовуються для вирішення морфологічної омонімії або для виявлення помилок в тексті відповідно.

1.5. Висновки до розділу

Результати, одержані в даному розділі, полягають в наступному:

1. Проведено аналіз відкритих та спеціалізованих джерел одержання інформації для вдосконалення комп'ютерних систем у результаті якого визначено шляхи збору та аналізу текстових даних, їх інтерпретації та

подальшого прийняття рішень щодо модернізації складових компонентів або системи в цілому.

2. Проаналізовано сучасний стан розвитку методів і засобів комп'ютерної лінгвістики на різних рівнях представлення текстової інформації та обґрунтовано необхідність застосування комплексного підходу щодо виявлення та аналізу тональності текстової інформації на синтаксичному, лексичному та семантичному рівнях.

3. Досліджено прикладні аспекти застосування комп'ютерної лінгвістики, що дало змогу визначити та обґрунтувати методи добування та аналізу відгуків користувачів про властивості комп'ютерних систем і її поведінки на основі аспектного підходу.

4. Проведено аналіз підходів до побудови лінгвістичних процесорів для вирішення задачі виявлення та оцінювання тональності текстової інформації, визначено їх переваги і недоліки, що дало змогу обґрунтувати застосування одночасно rule based і machine learning підходів.

РОЗДІЛ 2
 ПОБУДОВА
 МОДЕЛЕЙ ТА
 ОБГРУНТУВАННЯ
 МЕТОДІВ І МЕТРИК
 ОЦІНЮВАННЯ
 ТОНАЛЬНОСТІ
 ТЕКСТОВОЇ
 ІНФОРМАЦІЇ ПРИ
 ВДОСКОНАЛЕННІ
 КОМП'ЮТЕРНИХ
 СИСТЕМ

2.1. Обґрунтування та
 формалізація основних понять при
 використанні аспектного підходу
 виявлення та оцінювання тональності
 текстової інформації

В основі аспектного підходу при визначенні тональності текстової інформації лежить поняття думки, що представляє опис враження користувача від властивості або комп'ютерної системи як єдиного цілого. Формально, опис, який включає аспект, можна представити у вигляді квантилю

$$(e_i, a_{ij}, s_{ijkl}, h_k, t_l), \quad (2.1)$$

де e_i – сутність, яка виражена у відгуку користувача;

a_{ij} – j -ий аспект i -ої сутності;

s_{ijkl} – тональність j -ого аспекту i -ої сутності;

h_k – користувач, який висловив думку про аспект або сутність;

t_l – час (дата), коли була висловлена думка користувача h_k .

Тональність s_{ijkl} може приймати значення за набору («позитивний», «негативний», «нейтральний»).

У випадку, коли у відгуку про комп'ютерну систему вказується думка про усю систему, то оперують із спеціальним аспектом, який позначений, як загальний. В загальному випадку, пара (e_i, a_{ij}) формують ціль відгуку.

Правила, які використовуються при аналізі тональності текстової інформації визначаються наступними пунктами:

1. Обов'язкове представлення думки користувача у вигляді квантилю (2.1) з однозначним зв'язком між його компонентами. Тональність відгуку s_{ijkl} повинна бути висловлена власником h_k про аспект a_{ij} сутності e_i в конкретний момент часу t_l .

2. Важливість наявності усіх компонентів $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$, оскільки за відсутності одного з них актуальність відгуку втрачає сенс. Наприклад, якщо у квантилі відсутня інформація про час висловлення думки, то актуальність висловлювання більше, ніж рік перед тим, і сьогодні є різними, оскільки комп'ютерна система за цей час могла зазнати значних змін. Важливим також є інформація про стейкхолдера думки, оскільки допомагає визначити цільові групи користувачів.

3. Формальне представлення (2.1) охоплює більшість, але не все можливе семантичне навантаження висловленої думки. Для прикладу, висловлювання: «Ця комп'ютерна система не зручна для людей з вадами слуху». Таке висловлювання оцінює комп'ютерну систему в цілому і вказує, що вона не враховує особливостей людей з вадами слуху, однак вона може бути зручною для звичайної людини. Необхідно відмітити, що сутність складається з компонентів, кожен з яких володіє певними властивостями. Таку сукупність прийнято представляти у вигляді ієрархічної або мережної моделі даних про об'єкт. Для прикладу, «Вартість обслуговування крокового двигуна комп'ютерної системи є високою». Таке твердження не вказує на те, що обслуговування всієї комп'ютерної системи є дорого вартісним, а лише дорогим є обслуговування її частини – крокового двигуна. В залежності від задачі оцінювання тональності текстової інформації, можна говорити також про атрибути крокового двигуна, зокрема про його

якість, ціну і т.д. Тобто формула (2.1) може бути розширена шляхом застосування підходу вкладеності сутностей.

4. Означення (2.1) забезпечує базис для перетворення неструктурованого тексту в структуровані дані. Представлення думки у вигляді квантилю дозволяє спроектувати структуру бази даних у вигляді реляційних відношень, що забезпечує ефективність застосування операцій над даними і їх зберігання у системах керування базами даних.

5. Означення (2.1) можна використовувати лише у випадку регулярного виразу, що не містить порівняння об'єктів. У випадку речень, що містить порівняння об'єктів необхідно проводити трансформацію формального визначення думок таким чином, щоб встановити об'єкти порівняння або їх властивості, а після цього розробити методи їх оцінювання.

В загальному випадку, модель сутностей, які виявлені у висловлюваннях користувачів про комп'ютерні системи, або їх компоненти, можна інтерпретувати наступним чином: деяка сутність e_i представляє собою множину аспектів $A_i = \{a_{i1}, a_{i2}, \dots, a_{in}\}$, при цьому сутність e_i може бути виражена будь-яким набором її складових або компонентів – $\{ee_{i1}, ee_{i2}, \dots, ee_{is}\}$. Кожен аспект $a_{ij} \in A_i$ сутності e_i може бути виражений будь-яким кінцевим набором її аспектів – $\{ae_{ij1}, ae_{ij2}, \dots, ee_{ijm}\}$.

Модель документу, або модель відгуку feedback містить висловлювання, що формує множину сутностей $\{e_1, e_2, \dots, e_r\}$ і підмножину аспектів з множини відгуків, висловлених користувачами $\{h_1, h_2, \dots, h_p\}$ у деякий фіксований момент часу. У кінцевому випадку, формується множина висловлювань у документі feedback, аналіз тональності якого передбачає розв'язання шести основних задач:

Задача 1. Суть задачі полягає у виявленні і кластеризації усіх сутностей. Кожен кластер позначається узагальненою сутністю e_i .

Задача 2. Наступний крок передбачає виявлення усіх аспектів сутностей і їх кластеризацію. Кожен аспектний кластер сутності e_i відображається унікальним аспектом a_{ij} .

Задача 3. Суть задачі полягає у виявленні і категоризації користувачів, які висловили відгук, з подальшим формування наборів їх ознак.

Задача 4. Виявлення часу, коли були висловлені думки користувачів і приведення дати і часу до єдиного формату.

Задача 5. Класифікація аспектів a_{ij} за категоріями «позитивний», «негативний», «нейтральний» та формування ранжованих списків у визначених категоріях.

Задача 6. Представлення виявлених сутностей, аспектів, користувачів, часу і тональності у вигляді квантилю $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$.

Відгуки про комп'ютерну систему представляють собою короткі тексти, які значно менші, ніж цілі документи, і складаються з кількох речень. Застосування синтаксичного і семантичного розбору на рівні речення у відгуках користувачів дає змогу частково визначити думку про той чи інший об'єкт, але не забезпечує цілісної картини у випадку взаємозалежності речень між собою.

Варто відмітити, що використання аспектів, або так званих настроїв користувачів залежать від контексту вживання аспектів. Класифікація ж на рівні документу або речення шляхом застосування частотних методів або шаблонів є недостатніми підходами, які в повній мірі визначають думки з приводу властивостей об'єкта, або самого об'єкта.

Навіть у випадку припущення, що кожен документ оцінює одну цілісну комп'ютерну систему в позитивному контексті, не означає, що автор позитивно думає про усі властивості системи. Аналогічно і до негативного контексту щодо певного атрибуту системи. Тому, більш ефективним підходом до побудови алгоритму визначення тональності текстової інформації для вдосконалення комп'ютерних систем є використання аспектного підходу.

При вираженні аспектів деякого об'єкту, в даному випадку комп'ютерної системи або її компонентів, найбільш часто вживаними є іменники або групи іменників. У випадку використання групи іменників, вважається, що до її складу повинно входити не більше, ніж три або чотири слова. Якщо у групі виділити окремі слова, які представляють аспектні терміни об'єкту, то в подальшому можлива неузгодженість із судженнями експертів.

У загальному випадку існує чотири основні підходи щодо визначення аспектів у тексті:

- методи частотного аналізу іменників або груп іменників;
- методи побудови відношень між висловлюваннями порівнянності та аспектними термінами;
- методи машинного навчання з учителем;
- методи на базі статистичних тематичних моделей.

2.2. Добування аспектів на основі частотних характеристик слів

При добуванні термінів кандидатів у аспекти велику роль відіграє частота їх вживання в аналізованій колекції текстів. Частотні методи працюють доволі ефективно при описі простих ознак об'єктів, в тому числі і комп'ютерних систем. Однак, як показує практика, серед множини іменників існує досить багато таких слів, які не відображають аспекти загальнолітературної лексики, або погано піддаються розпізнаванню слова-іменники, які мало зустрічаються у тексті.

На практиці та у наукових публікаціях широкого застосування набув метод інформаційного пошуку за ознаками – TF-IDF. Даний метод дозволяє враховувати частоту вживання термінів як на рівні документу, так і на рівні абзацу. Підхід TF-IDF (рис. 2.1) використовують для визначення аспекту термінів порівняння частот іменних груп у колекції відгуків з відповідними

частотами цих груп у контрастній групі Національного британського корпусу.

У випадку, коли аспект термінів добуває не тільки окремо взяті іменники, а групи іменників, то необхідно застосовувати додаткові ознаки для більш точного визначення довжини групи слів. Найбільш часто використовують так звані контекстні ознаки, які оцінюють частоту вживання словосполучення з частотою контексту. Такі ознаки дозволяють визначити межі груп іменників. Для цього можна використати міру FRL (2.2) або LR (2.3)

$$FLR(a) = f(a) \cdot R(a) \quad (2.2)$$

$$LR(a) = \sqrt{l(a) \cdot r(a)} \quad (2.3)$$

де $f(a)$ – частота вживання аспектного терміну;

$l(a)$ – кількість різних слів, що знаходяться зліва від a ;

$r(a)$ – кількість різних слів, що знаходяться справа від a .



Рис. 2.1. Метод TF-IDF

Наступний крок полягає у відборі груп іменників з мірою, що більша за середню міру для словосполучень. Таким чином, дана метрика в першу чергу формує групи іменників, які мають велику різноманітність слів на

своїх границях, і показує, що термін *a*, який аналізується не є фрагментом довшого словосполучення.

Інший підхід до добування аспектних термінів або груп іменників, базується на використанні ознаки *C-value*, що дає змогу знизити вагу даного слова або словосполучення, якщо воно входить у частотне словосполучення більшої довжини. Використання ознаки *C-value* при аналізі відгуків про комп'ютерні системи є важливим моментом, оскільки дає змогу керувати пріоритетом важливих і не важливих властивостей комп'ютерних систем.

Це дозволяє врахувати той факт, що словосполучення, до складу якого входить більше слів, може бути включено у список кандидатів на аспектні терміни, а поточне словосполучення є його частиною (фрагментом).

Інший підхід щодо ідентифікації термінів-аспектів передбачає, що такими термінами можуть бути лише ті групи іменників, які вживаються у вигляді присудку, дієслів або групи, що складає підмет речення. Даний підхід базується на застосуванні методів синтаксичного розбору речень і може використовуватись у підходах *rule-based*.

Алгоритм, описаний у [17], працює за принципом виключення потенційних аспектних термінів. Передбачається, що ті слова, які не зустрічаються досить часто у заданих шаблонах, що позначають частину-ціле. При цьому на основі пошуку в мережі Інтернет обчислюється показник РМІ (*Pointwise Mutual Information*) вживання потенційного аспектного терміну з цільовим об'єктом. Для прикладу, в англійській мові для цифрових камер перевіряється вживання аспектних термінів у шаблонах типу «*of camera*», «*camera has*».

Окрім цього, при визначенні аспектних термінів може застосовуватись так звана ієрархія тезауруса *WordNet*. Вона призначена для виявлення назв об'єктів або їх частин, а також словоутворюючих суфіксів типу (*-iness*, *-ity*) для англійської мови.

Ієрархія тезауруса дає змогу визначати аспектні терміни у технічних областях, зокрема і при аналізі відгуків про комп'ютерні системи. Однак

WordNet є не ефективним при аналізі текстів щодо відгуків про фільми, програмне забезпечення, сферу обслуговування. При цьому такий підхід дороговартісний і розтягнутий у часі, оскільки потребує надійного каналу для забезпечення зв'язку з мережею Інтернет.

Суть методів частотного аналізу вживання певних термінів і словосполучень дає змогу визначити найбільш важливі, з точки зору авторів відгуків, властивості комп'ютерних систем або враження від комп'ютерної системи в цілому. Однак він не дає змоги провести аналіз семантичного навантаження у висловлюваннях.

2.3. Визначення аспектних термінів на основі фраз-порівнянь

2.3.1. Ітеративні методи для виявлення термінів-аспектів

З практики та наукових публікацій відомо, що термін-аспект обов'язково повинен входити у шаблони порівняльних речень або речень, у яких дається оцінка деякому об'єкту чи властивості.

Для визначення зв'язку між аспектом і словами-порівняння використовується синтаксичний аналізатор. Відношення між аспектом та оціночним словом добуваються на основі заданих шляхів синтаксичної залежності. Так, наприклад, у реченні «This system is not an ideal» слово system і слово ideal будуть розміченими відповідно як аспект і оціночне слово, оскільки між ними існує шлях у синтаксичному дереві «NN – nsubj – VB – dobj – NN» (іменник-підмет – присудок – дієслівна група – пряме доповнення – іменник).

При добуванні аспектних термінів на основі їх взаємозв'язку із словами, що виражають оцінювання, часто використовуються ітеративні методи (bootstrapping).

В якості вхідної множини початкових значень можна використати частотні групи іменників, які представляють аспекти, або задати їх вручну.

Основні великі аспекти термінів (частотні слова та групи іменників) використовуються для виявлення асоціативних правил, тобто, шаблонів, які мають різні аспекти і зазвичай пов'язані із словами, що виражають певну оцінку властивості чи комп'ютерної системи в цілому.

Після формування таких правил отримують менш частотні аспекти термінів, тобто такі групи іменників, які з'являються саме в таких шаблонах з оціночними словами.

На основі описаного вище підходу, для аналізу тональності відгуків користувачів комп'ютерних систем, можна використовувати підхід *double propagation* (подвійне розповсюдження) для утворення нових аспекти термінів і розширення довідників слів для вираження оцінки іменників.

В якості вхідного набору задається множина, сформована на основі словника з невеликою кількістю слів-оцінювачів і синтаксичні шаблони, які містять комбінації оціночних слів і термінів-аспектів. У результаті, виходячи з відомого виразу або слова для оцінювання, можна виділити аспекти, і навпаки, з аспектного терміну виділити слова для оцінювання.

Очищення одержаної множини аспектів реалізується на основі ряду правил. Наприклад, пропонується, що в одному фрагменті речення без розділового знаку кома, міститься лише один аспекти термін, а інший кандидат в аспекти повинен бути видалений. Видаленню підлягає менш частотний термін, який існує у колекції.

На практиці оцінювання цього методу відбувалось у п'яти галузях, у результаті чого одержано *F*-міру на рівні 85%, а кількість відгуків, які були проаналізованими становила близько 65000.

Для врахування важливості аспекти термінів можна ввести ще два критерії. Перший критерій відповідає за різноманітність слів для оцінювання кандидата в аспекти при оцінюванні комп'ютерної системи – різноманітність свідчить про важливість аспекти, тобто чим більшою кількістю слів описано термін, тим більша його вага. Другий критерій

стосується пошуку відношення аспектного терміну з певним об'єктом на основі заданих шаблонів.

Наприклад, при описі комп'ютерної системи англійською мовою можна знайти такі фрази: *the camera of the system* (камера системи), *the system has a great camera* (система має чудову камеру), що свідчить про відношення частина-ціле між камерою і комп'ютерною системою. Якщо слово одночасно вживається і з оціночним словом, і у відношенні із заданим об'єктом, то це означає, що знайдено важливий аспект, якому присвоюється велике значення ваги.

При ітераційному пошуку аспектних термінів використовується деяка початкова множина аспектів, які можна доповнювати на основі:

- врахування міри взаємної інформації щодо розташування аспекта-кандидата в одних і тих же реченнях, що й аспекти з початкової множини і врахуванні частоти вживання аспекта-кандидата;

- при доповненні множини аспектів варто проводити очищення надлишкових аспектів.

Для технічної сфери та інформаційних технологій кількість аспектів може досягати 200 одиниць.

2.3.2. Методи машинного навчання в задачах виявлення аспектів

Існує два способи використання методів машинного навчання для виявлення аспектів у відгуках про комп'ютерні системи:

- методи, що базуються на попередньому формуванні набору аспектних термінів у деякій предметній області з подальшим навчанням моделі, яка використовує ознаки аспектів, які розглядались до цього;

- методи засновані на маркуванні (розмітці) послідовності слів у відгуках (розмітка аспектних термінів, слів оцінювання).

При визначенні аспектних термінів, окрім характеристики частоти вживання аспектів-кандидатів у відгуках про комп'ютерні системи, використовуються порівняння кандидатів із заголовками словникових статей у Вікіпедії. Семантична близькість кандидатів розраховується на основі сукупності посилань відповідної статті у Вікіпедії, а також величини асоціації кандидата в аспекти з назвою сутності при пошуку і Інтернеті. У результаті практичного застосування цього методу для кількох об'єктів F -міра оцінюється у 72,7%.

Як приклад, наведемо виявлення аспектних термінів на основі аналізу відгуків користувачів про ноутбук, які були написані кирилицею. При цьому в якості набору ознак визначимо множину вхідних аспектних термінів у вигляді окремих іменників:

- частота в колекції відгуків;
- близькість до оціночних слів (вікно розміром p), яке в даному випадку розглядається не як близькість до оціночних слів у колекції таких слів, а близькість до слів хороший/поганий при видачі результатів пошуку відповідними системами;
- ознака відмінності, що обчислюється як відносна частота слова у порівнянні з контрастною колекцією слів;
- ознака TF-IDF;
- міра взаємної інформації PMI, що враховує спільну частоту вживання між іменниками кандидатами в аспекти і визначеним типом системи (ноутбук).

На основі використання різних варіантів для кожної з мір одержано результати, які описують двадцять три ознаки ноутбука, а результат виявлення аспектів коливається в межах від 76% до 86% для різних областей застосування.

Проте, найбільш популярними в області виявлення аспектних термінів на основі методів машинного навчання є підходи, що базуються на послідовній розмітці аспектів у корпусів документів. Дод розмічених даних

застосовують методи НММ (Hidden Markov Models) та CRF (Conditional Random Fields). У якості ознак використовуються такі критерії як слова, частини мови, синтаксичні залежності, відстані, речення, що містять слова для оцінювання та ін. Ці ж моделі можуть використовуватись і для спільного виявлення аспектів і оціночної лексики. Методи, що використовують машинне навчання можуть виявляти і низькочастотні аспекти, однак потребують розмітки даних, що є доволі трудозатратним процесом при застосуванні послідовних методів машинного навчання.

2.3.3. Тематичні моделі для виявлення аспектних термінів

Виявлення аспектів у відгуках користувачів при аналізі відгуків про комп'ютерну систему або її властивості можна на основі імовірнісних тематичних моделей. У цьому випадку передбачається, що кожен відгук сформований з набору прихованих тем, а кожна тема представляє собою імовірний розподіл слів.

Зазвичай, на практиці використовують два типи тематичних моделей:

- pLSA (probabilistic Latent Semantic Analysis);
- LDA (Latent Dirichlet Allocation).

У результаті застосування тематичних моделей до колекції текстів утворюється сукупність тем, кожна з яких представляє собою список слів з імовірностями їх приналежності до тієї чи іншої теми.

При визначенні тональності текстової інформації припускають, що при застосуванні обраної тематичної моделі, виділені теми будуть містити аспектні терміни, які відносяться до певної аспектною категорії.

Для виявлення аспектів потрібно модифікувати базові тематичні моделі шляхом формування двох окремих тем: слова, що виражають оцінку об'єкту та аспекти самі по собі. При успішному застосуванні таких моделей відбувається одночасно дві операції: виявлення аспектів і їх групування за узагальнюючими категоріями.

Застосування класичної моделі LDA, яка будується на основі даних про взаємне вживання слів в одних і тих же текстах не є ефективною, оскільки у множині різних відгуків, наприклад, про комп'ютерні системи, може містити один і той самий набір аспектів.

Однією з найбільш відомих тематичних моделей, що побудована на основі LDA і дозволяє виявляти аспекти комп'ютерних систем, застосовується глобальна модель добування іменованих сутностей, а для виявлення аспектів термінів використовують вікно із слів або речень. Вживання слів таких фрагментів дозволяють визначити аспекти, але вони не розрізняють аспекти та слова, що оцінюють деякий об'єкт.

Для прикладу, у темі «Обслуговування» визначено такі аспекти як: «staff, friendly, helpful, service, desk, concierge, excellent, extremely, hotel, great, reception, English, pleasant, help».

Застосування LDA моделі у чистому вигляді є неефективним засобом виявлення аспектів, а гібридні моделі, зокрема комбінація MaxEnt-LDA, у якій відбувається спільне виявлення і добування аспектичних та оціночних слів на основі синтаксичних ознак, дають змогу розділити їх.

Метод максимум ентропії можна застосовувати при підборі параметрів на розмічених даних.

Однак застосування тематичних моделей при добуванні аспектів характеризується наступними проблемами:

- виникає необхідність використання великих об'ємів даних і дуже чутливе налаштування параметрів моделі для одержання достовірних результатів;
- не стабільність та нестійкість результатів виявлення аспектів;
- тематичні моделі доволі просто визначають частотні аспекти, які ефективно визначаються за допомогою інших методів.

2.4. Методи аналізу тональності текстової інформації при вдосконаленні комп'ютерних систем

Для аналізу текстової інформації при вирішенні задачі вдосконалення комп'ютерних систем, можна застосовувати, як і у випадку аналізу тональності документів і речень, методи машинного навчання та інженерно-лінгвістичні методи.

Основне завдання, яке необхідно при цьому вирішити, полягає у визначенні оцінок аспектів та діапазону дії кожного речення, що містить слова для оцінювання аспектів. Одним з найбільш широко використовуваних підходів є побудова та аналіз синтаксичних конструкцій речень у вигляді дерев залежностей.

2.4.1. Методи машинного навчання для аналізу тональності аспектів

При використанні методів машинного навчання формуються словники термінів, що оцінюють аспекти, та самі аспекти. Після цього будується класифікатор, який визначає оцінки для кожного аспекту на основі оціночних слів. Це означає, що класифікатор повинен визначити до якого аспектного терміну відноситься слово, що містить оцінювання об'єкту. У випадку речення з великою кількістю слів, значно зростає обчислювальна складність, оскільки воно може містити багато слів з оцінюванням комп'ютерної системи або її складових, а також багато аспектів (хороша камера, але поганий звук).

В якості ознак аспектів і термінів для оцінювання об'єктів використовуються:

- ознаки розташування: відстань між аспектним терміном і словом-оцінювачем, кількість аспектів і спеціальних слів у реченні, довжина

речення, знаки пунктуації, наявність одних аспектів серед інших, порядок розташування аспектів та слів-оцінювачів;

- лексичні ознаки: набір слів між аспектним терміном і оціночним словом, наявність спеціальних часток та ін.;

- частини мови для слів, що оцінюють об'єкт комп'ютерної системи та аспектного терміну, набір тегів частин мови між аспектом та словом-оцінювачем, частини мови сусідніх слів;

- ознаки, побудовані на синтаксичній структурі: набір тегів на шляху між аспектним терміном і словом, що його оцінює, близькість на синтаксичному дереві.

Всі ці ознаки є важливими при визначенні пар «аспектний термін-слово, що його оцінює», оскільки при експериментах встановлено, що міра точності складає близько 83%. Базовий рівень для порівняння, передбачав, що слово-оцінювач приписується найближчому аспекту і становив 77%.

2.4.2. Лінгвістично-інженерні методи оцінювання аспектів

Лінгвістично-інженерні методи оцінювання аспектів передбачають наявність на момент класифікації наступних даних:

- назви сутностей і їх аспектів;
- існує словник слів для оцінювання аспектів і фраз, а також правила їх перетворення в залежності від контексту і адитивного правила.

Аналіз тональності відбувається по кожному реченні і передбачає ряд етапів попереднього опрацювання слів.

Спочатку виконується маркування відомих аспектних термінів на рівні речення та слів, які їх оцінюють. Для оціночних слів відома оцінка тональності, яка в найпростішому вигляді лежить в діапазоні від 1 до -1. До слів, які оцінюють аспекти, можливе застосування операторів, що здатні змінити їх тональність на протилежну. Як варіант можливе застосування часток «не».

Наступний крок полягає у визначенні структури речення з метою можливої модифікації базових оцінок, зокрема важливість опрацювання частки «не» і слів по-типу «однак», «проте» і т.д. Якщо у другій частині речення не виявлено слів-оцінювачів, але наявною є частка «не», то цій частині речення повинна бути поставлена оцінка, обернена до оцінки, яка поставлено у першій частині речення.

Після виконання двох попередніх етапів, необхідно провести агрегацію оцінок за кожною аспектною категорією.

Інший метод, який належить до лінгвістично-інженерного підходу, полягає у послідовному оцінюванні тональності у кожному реченні у відгуку про комп'ютерні системи. Формально постановку такої задачі можна представити наступним чином: нехай у реченні S міститься набір аспектних термінів a_1, a_2, \dots, a_n та виразів, що їх оцінюють – sw_1, sw_2, \dots, sw_n , для яких оцінки із словника модифіковані згідно контексту та операторів. Тоді оцінки аспектних термінів обчислюються за формулою

$$\text{score}(a_i, S) = \frac{sw_j so}{\text{dist}(sw_j, a_i)}, \quad (2.4)$$

де sw_j – слово, для оцінювання аспектів;

$sw_j so$ – кількісна оцінка тональності sw_j ;

$\text{dist}(sw_j, a_i)$ – відстань між словами, що оцінюють аспект, та самим аспектом.

Таким чином, для кожного аспектного терміну у реченні визначаються всі оцінки, які наявні у ньому, однак їх вага спадає в залежності від відстані між аспектом і словом. Якщо результат формули (2.4) більший за нуль, то і оцінка позитивна. У випадку, коли результат від'ємний – відгук негативний, а якщо рівний нулю – нейтральний.

Таким чином, найбільш ефективним способом для оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем

є підхід лінгвістично-інженерних методів, який дозволяє проводити класифікацію за трьома класами: позитивний, негативний та нейтральний.

2.5. Методи і метрики оцінювання якості результатів класифікації

При оцінюванні якості класифікації текстової інформації можуть використовуватись різні метрики. Найбільш використовуваними є наступні:

- accuracy (точність);
- precision (точність) та Recall (повнота);
- confusion Matrix (матриця неточностей);
- F-міра.

Проведемо аналіз та обґрунтуємо доцільність застосування метрик при оцінюванні тональності у відгуках користувачів для вдосконалення комп'ютерних систем.

2.5.1. Метрика Accuracy

Найпростішою метрикою при оцінюванні тональності текстової інформації є міра, що відображає правильність класифікації відгуків за аспектами комп'ютерної системи. Метрика Accuracy обчислюється як відношення кількості правильно класифікованих відгуків до загальної кількості відгуків у навчальній вибірці

$$\text{Accuracy} = \frac{\text{Right}}{\text{Total}}, \quad (2.5)$$

де Right – кількість відгуків за якими класифікатор правильно визначив приналежність до класу позитивних, негативних чи нейтральних відгуків;

Total – загальна кількість відгуків, тобто об'єм навчальної вибірки.

Особливістю використання метрики (2.5) є те, що усім відгукам про комп'ютерну систему присвоюється однакова вага. Це може призвести до ситуації, коли будуть значні відхилення у випадку розподілу відгуків у навчальній вибірці щодо приналежності до одного або кількох класів. А це в свою чергу означає, що більшість відгуків будуть адекватно класифіковані за пріоритетними класами навчальної вибірки. На практиці використання метрики Ассурасу призводить до того, що її значення буде на рівні 70%, однак реально відгуки не відображають реального стану речей. Для уникнення такого негативного впливу, класифікатор необхідно навчати на спеціально визначеному наборі даних, який є збалансованим корпусом документів.

До недоліків метрики Ассурасу належить те, що дана міра не відображає інформації про відносну частоту відгуку (документу). В окремих випадках, така характеристика є визначальною при прийнятті правильного рішення.

2.5.2. Метрика точності і повноти

Точність і повнота представляють собою метрику, яка застосовується на практиці при оцінюванні більшості алгоритмів виявлення (добування) інформації, в тому числі і аспектів. У деяких випадках, зокрема при оцінюванні якості рекомендаційних систем, ця метрика використовується само по собі, а у деяких, наприклад при оцінюванні класифікаторів, вона становить базис для інших метрик, зокрема, F-міри та R-Precision.

Під точністю системи у межах класу розуміють частку документів, які дійсно належать до деякого класу відносно сукупності всіх документів, які класифікатор позначив міткою даного класу.

Повнота системи представляє собою відсоток знайдених класифікатором відгуків, які належать до даного класу відносно набору усіх відгуків цього класу у тестовому наборі даних.

Значення точності і повноти можна розрахувати за таблицею контингентності, приклад якої наведено у табл. 2.1.

Таблиця 2.1

Матриця контингентності

Клас _{<i>i</i>}		Експертна оцінка	
Оцінка системи	Позитивна	<i>TP</i>	<i>FP</i>
	Негативна	<i>FN</i>	<i>TN</i>

Табл. 2.1 містить інформацію про те, яку кількість відгуків класифікатор правильно і неправильно відніс до визначеного класу, зокрема:

TP – істинно-позитивне значення;

TN – істинно-негативне значення;

FP – хибно-позитивне значення;

FN – хибно-негативне значення.

У даному випадку, точність і повноту можна визначити шляхом застосування формул (2.6) і (2.7) відповідно

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2.6)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2.7)$$

У якості прикладу застосування точності і повноти, розглянемо аналіз відгуків користувачів про комп'ютерну систему. Припустимо, що існує

тестова вибірка з кількістю відгуків 100, серед яких 60 відгуків позитивних і 40 негативних. Провівши аналіз, класифікатор позначив лише 20 негативних відгуків, при цьому 10 є істинно-негативними, а ще 10 визначено як позитивні. У результаті одержуємо 10 істинно-позитивних рішень, 10 хибно-позитивних та 30 хибно-негативних. Тоді для класу негативних відгуків одержуємо точність $Precision = \frac{10}{10+10} = \frac{1}{2}$, або точність класифікації становить 50%.

Значення повноти, у наведеному прикладі, становить $Recall = \frac{10}{10+30} = \frac{1}{4}$, або це означає, що класифікатор знайшов лише 25% з усіх негативних відгуків.

2.5.3. Матриця контингентності

З практичної точки, зручніше оперувати значеннями точності і повноти за допомогою матриці контингентності (неточностей). У випадку, коли кількість класів (аспектів) є не великою, тобто становить до 100-150 одиниць, підхід confusion matrix забезпечує наглядність результатів аналізу даних класифікатором. Приклад матриці неточностей наведено на рис. 2.2.

Таблиця контингентності представляє собою квадратну матрицю розміром N на N елементів, де розмірність вказує на кількість класів. Стовпці цієї матриці заповнюють експертними оцінками, а рядки – результатами роботи класифікатора. При класифікації відгуку користувача про комп'ютерну систему (аспект), або його компонент, з тестової вибірки, проводиться збільшення значення, що знаходиться у комірці на перетині стрічки класу, який визначив класифікатор, і стовпця класу, до якого істинно належить відгук.

	0.91	0.96	0.94	0.75	1.00	0.83	0.85	0.97	1.00	0.86	1.00	0.79	1.00	0.75	1.00	1.00	0.96	0.90	0.81	0.89	0.94	0.98	0.86	0.89	0.94	0.92	0.96
0.80	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
0.95	1	94	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1.00	2	0	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.29	3	0	0	6	0	0	3	2	0	1	0	0	0	0	0	0	1	1	0	0	1	0	1	3	0	2	0
1.00	4	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.50	5	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	1
0.92	6	1	0	0	0	0	152	0	0	1	0	0	0	0	0	0	0	1	4	2	3	0	0	0	0	2	0
0.97	7	1	0	1	0	0	0	256	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	2	0
0.33	8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0
0.97	9	0	0	0	0	0	0	0	0	69	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
0.82	10	0	0	0	0	2	0	0	0	0	18	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
0.87	11	0	0	0	0	0	0	0	0	0	0	34	0	4	0	0	0	0	0	0	0	0	0	0	1	0	0
1.00	12	0	0	0	0	0	0	0	0	0	0	0	37	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.57	13	0	0	0	0	0	0	0	0	0	0	9	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0
0.63	14	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	3	0	0	0	0	0	0
0.50	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1	1	0	0
0.77	16	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	47	0	1	3	4	0	0	2	0	1	0
0.87	17	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	69	1	2	5	0	0	0	0	0	0
0.97	18	0	0	0	0	1	4	0	0	1	0	0	0	0	0	0	0	197	1	0	0	0	0	0	0	0	0
0.78	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	35	183	13	0	0	2	0	1	0	
0.97	20	0	0	0	0	10	3	0	1	0	0	0	0	0	0	0	0	4	702	0	0	0	0	0	0	6	0
0.93	21	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	56	0	2	0	0	
0.29	22	0	0	1	0	0	2	0	0	6	0	0	0	0	0	0	0	1	1	1	0	6	2	0	1	0	
0.91	23	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	3	6	0	0	115	0	0	0	
1.00	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	
0.93	25	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	4	5	0	0	0	1	196	0	
0.98	26	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	78	

Рис. 2.2. Приклад матриці неточностей (confusion matrix)

Виходячи з аналізу матриці, наведеної на рис. 2.2, видно, що класифікатор у більшості випадків правильно визначає аспекти комп'ютерної системи і діагональні елементи є явно вираженими. Однак при класифікації щодо приналежності аспектів до класів 3, 5, 8, 22 точність є доволі низькою.

При застосуванні confusion matrix, точність класифікації визначається відношенням значення відповідного діагонального елементу і суми значень у всій стрічці матриці.

Повноту ж можна визначити шляхом обчислення відношення значення діагонального елемента матриці і суми значень у відповідному стовпці.

У формальному вигляді при застосуванні матриці контингентності точність можна описати формулою

$$\text{Precision} = \frac{A_{c,c}}{\sum_{i=1}^n A_{c,i}}, \quad (2.8)$$

$A_{c,c}$ – значення, що міститься на діагоналі матриці при перетині стовпця c і стрічки c ;

$\sum_{i=1}^n A_{c,i}$ – сума значень на перетині стовпця c і стрічки i .

Загальна точність і повнота класифікатора обчислюється як середнє значення точності за усіма класами і у пакетах інтелектуального аналізу даних називається *macro-averaging*. З прикладу, наведеному на рис. 2.1 видно, що кількість класів становить 26, загальна точність класифікації 80%, а повнота – 91%.

2.5.4. Застосування F-міри

У випадку застосування метрик точності і повноти, інтерпретацію результатів можна трактувати наступним чином: чим вища точність і повнота, тим кращий результат. Однак на практиці, необхідно знаходити баланс між значеннями точності і повноти, оскільки неможливо одночасно досягти максимального значення цих метрик. Тому, необхідно застосовувати деяку проміжну метрику, яка б враховувала баланс між Precision та Recall. Однією з таких метрик є F-міра, що представляє собою середнє гармонічне між точністю і повнотою. У випадку, коли точність і повнота прямують до нуля, то значення F-міри також прямує до нуля. Формально дану метрику можна записати у наступному вигляді

$$F = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.9)$$

Згідно формули (2.9) точність і повнота володіють однаковим пріоритетом, тому зниження F-міри відбувається при одночасному зменшенні і повноти, і точності.

У випадку, коли один з параметрів (точність або повнота) є більш важливим при досягненні деякої мети, до F-міри можна застосувати вагові коефіцієнти і формула (2.9) перетвориться наступним чином

$$F = (\beta^2 + 1) \cdot \frac{\text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}}, \quad (2.10)$$

де β – ваговий коефіцієнт, область визначення якого належить діапазону $(0; \infty)$.

Якщо значення $0 < \beta < 1$, то пріоритет надається точності класифікації. У випадку, коли $\beta > 1$ – пріоритетом виступає повнота. При $\beta = 1$ – формула (2.7) зводиться до формули збалансованої F-міри. Збалансовану F-міру прийнято називати як F1-міра.

На рис. 2.3 наведено візуалізацію збалансованої F-міри.

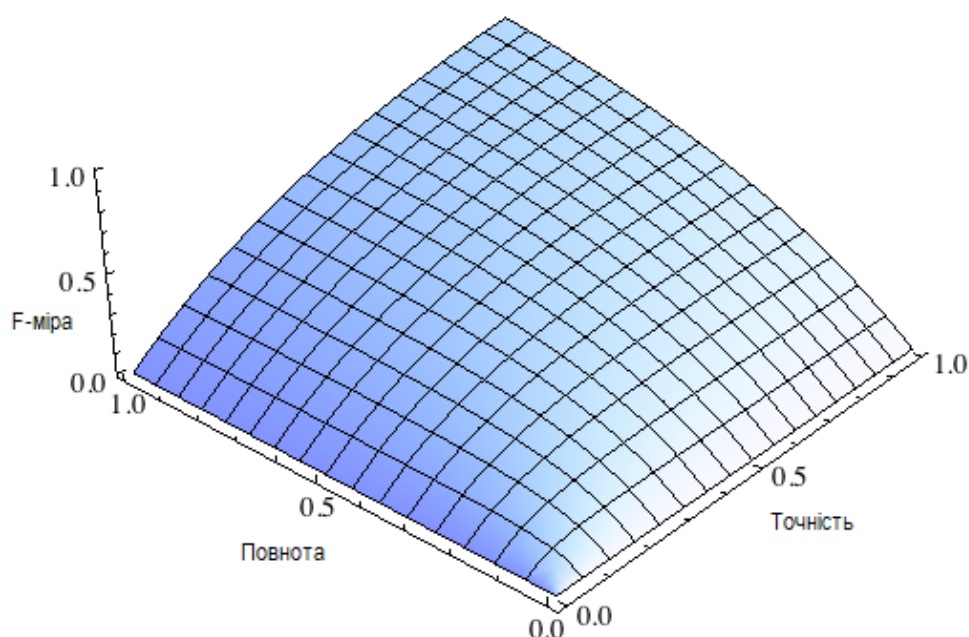


Рис. 2.3. Збалансована F-міра

Представлення F-міри з пріоритетом точності $\beta^2 = \frac{1}{4}$ наведено на рис. 2.4.

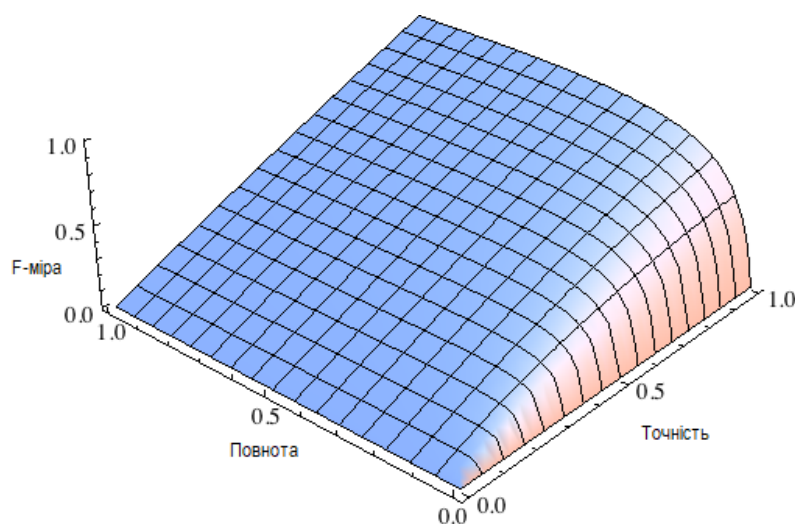


Рис. 2.4. F-міра з пріоритетом точності

На рис. 2.5 наведено приклад візуалізації F-міри для випадку, коли $\beta^2 = 2$.

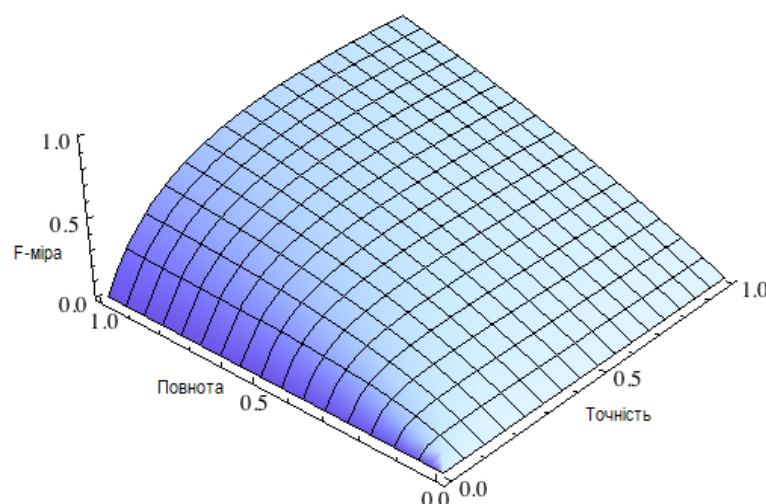


Рис. 2.5. F-міра з пріоритетом повноти

F-міра може ефективно використовуватись при валідації якості класифікатора, оскільки дає змогу кількісно, на основі одного значення, виразити дві інші найбільш важливі метрики – точність і повноту.

Керуючи значенням пріоритетів повноти і точності, можна оцінити зміни в алгоритмі та його прийнятність. При визначенні тональності текстової інформації щодо властивостей комп'ютерних систем пропонується скористатись саме F-мірою.

2.6. Висновки до розділу

1. Формалізовано представлення відгуків користувачів і задач, які при цьому необхідно розв'язати, у вигляді квантилю (сутність-аспект-тональність-користувач-час), що дало змогу врахувати позитивні і негативні властивості, реалізовані у комп'ютерних системах та в подальшому відобразити їх при вдосконаленні систем.

2. Обгрунтовано необхідність застосування частотних методів при виявленні аспектів комп'ютерних систем, що позначаються на враженні кінцевих користувачів і дають змогу виявити найбільш важливі властивості комп'ютерних систем.

3. Проведено аналіз методів добування аспектів на основі підходу побудови правил, машинного навчання і тематичного моделювання, у результаті якого обгрунтовано необхідність використання методів машинного навчання в комплексі з частотними методами синтаксичного аналізу при вирішенні задач вдосконалення комп'ютерних систем.

4. Обгрунтовано застосування лінгвістично-інженерного підходу при оцінюванні тональності текстової інформації щодо вдосконалення комп'ютерних систем, що дає змогу врахувати якість аспектів на основі слів, що їх оцінюють та відстані між аспектами і такими словами.

5. Проаналізовано методи і метрики оцінювання тональності текстової інформації і встановлено доцільність застосування F-міри, що враховує повноту і точність оцінювання тональності відгуків користувачів і дає змогу виявити найбільш проблемні властивості комп'ютерної системи.

РОЗДІЛ 3
СИСТЕМА
ВИЯВЛЕННЯ ТА
АНАЛІЗУ
ТОНАЛЬНОСТІ
ТЕКСТОВОЇ
ІНФОРМАЦІЇ ПРИ
ВДОСКОНАЛЕННІ
КОМП'ЮТЕРНИХ
СИСТЕМ

3.1. Функціональні вимоги до
системи виявлення та аналізу
тональності текстової інформації

Основні функціональні вимоги до системи виявлення та аналізу тональності текстової інформації повинні відображати процеси предметної області щодо вдосконалення комп'ютерних систем і базуватись на запропонованих та обґрунтованих методах і моделях аспектного семантичного аналізу.

Основними користувачами такої системи є наступні групи користувачів:

- інженери з вдосконалення комп'ютерних систем;
- аналітики комп'ютерних систем;
- керівники компанії-розробника комп'ютерної системи.

Для кожного з цих типів ролей необхідно визначити ключові функціональні вимоги та представити їх у графічному вигляді. У роботі пропонується скористатись нотаціями діаграми прецедентів, що входить до стандартного набору UML діаграм.

Так, основними функціональними можливостями системи підтримки процесів виявлення та аналізу тональності текстової інформації при використанні інженером з вдосконалення комп'ютерних систем є:

- формування набору властивостей комп'ютерних систем, які необхідно врахувати при їх вдосконаленні;

- аналіз відгуків користувачів про аспекти системи;
- розробка плану інтеграції нових властивостей;
- перевірка відповідності реалізованих властивостей щодо плану модернізації.

На рис. 3.1 наведено сценарій виконання функціональних можливостей інженера з вдосконалення комп'ютерних систем.

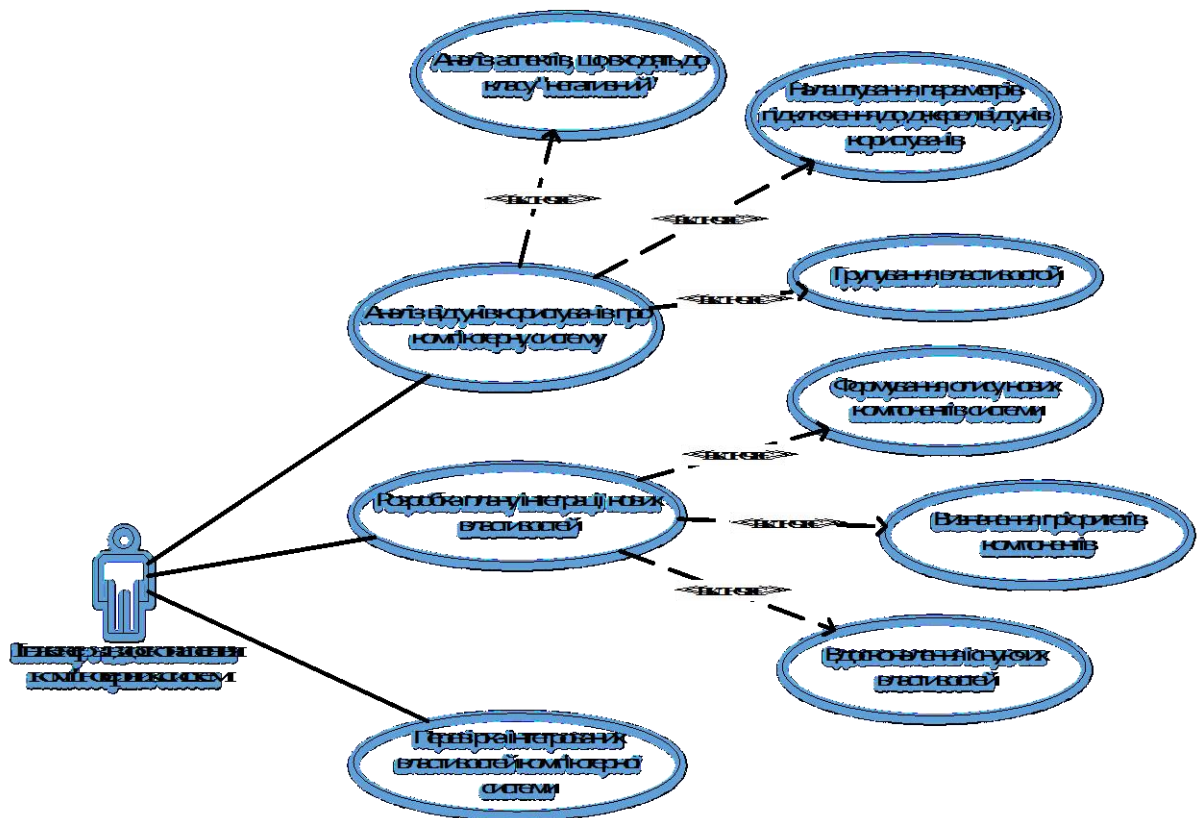


Рис. 3.1. Use case діаграма «Інженер з вдосконалення комп'ютерних систем»

Зона відповідальності інженера з вдосконалення та розвитку комп'ютерної системи визначається аналізом реалізованих властивостей комп'ютерної системи, постійним моніторингом функціонального стану системи. На основі одержаних результатів оцінювання якості аспектів та рекомендацій керівництва компанії, інженер розробляє план інтеграції або внесення змін у комп'ютерну систему, а після цього проводить тестування і перевірку на відповідність.

Аналітики комп'ютерних систем є ключовими особами при впровадженні інтелектуальних методів аналізу тональності текстової інформації при вдосконаленні комп'ютерних систем. Основні функції, які необхідно забезпечити аналітикам наведені на рис. 3.2.

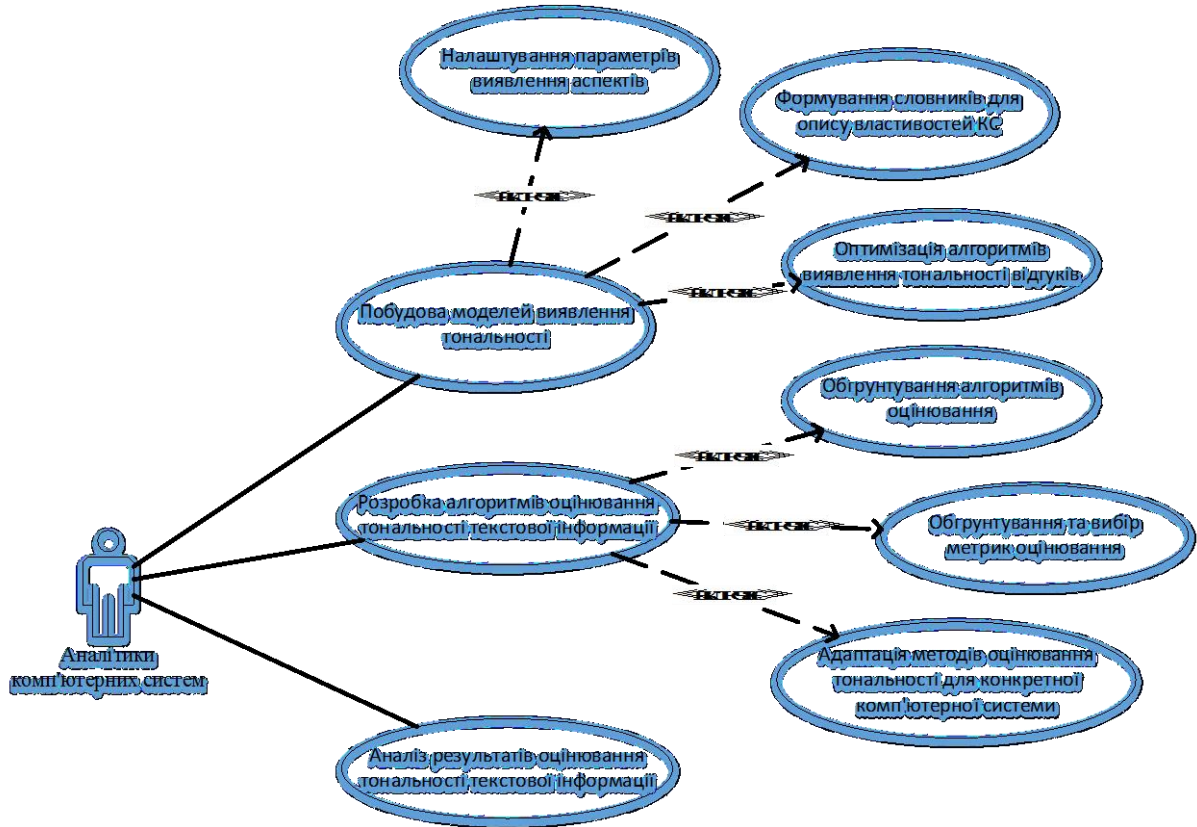


Рис. 3.2. Діаграма варіантів використання «Аналітики комп'ютерних систем»

Аналітики комп'ютерних систем забезпечують розв'язок наступних задач:

- обґрунтовують і будують моделі для виявлення тональності текстової інформації у відгуках користувачів комп'ютерних систем;
- забезпечують визначення словникових термінів щодо аспектів та слів, які оцінюють властивості комп'ютерних систем;
- обґрунтовують алгоритми виявлення та оцінювання тональності текстової інформації;
- розробляють шаблони представлення речень, навчальні і тестові набори даних

- виконують розмітку тексту
- обґрунтовують метрики точності виявлення аспектів та оцінювання тональності текстової інформації
- проводять адаптацію параметрів визначених моделей для конкретної сфери застосування.

Як результат аналізу діаграми, наведеної на рис. 3.2., функції аналітика є найбільш важливими при впровадженні інтелектуальних методів дослідження текстової інформації при вдосконаленні комп'ютерних систем.

Для затвердження плану вдосконалення та визначення стратегії розвитку комп'ютерної системи у засобі виявлення та аналізу тональності текстової інформації визначено ще одну роль – керівник компанії-розробника. На рис. 3.3 наведено основні функціональні можливості керівника компанії, що проводить вдосконалення комп'ютерної системи.

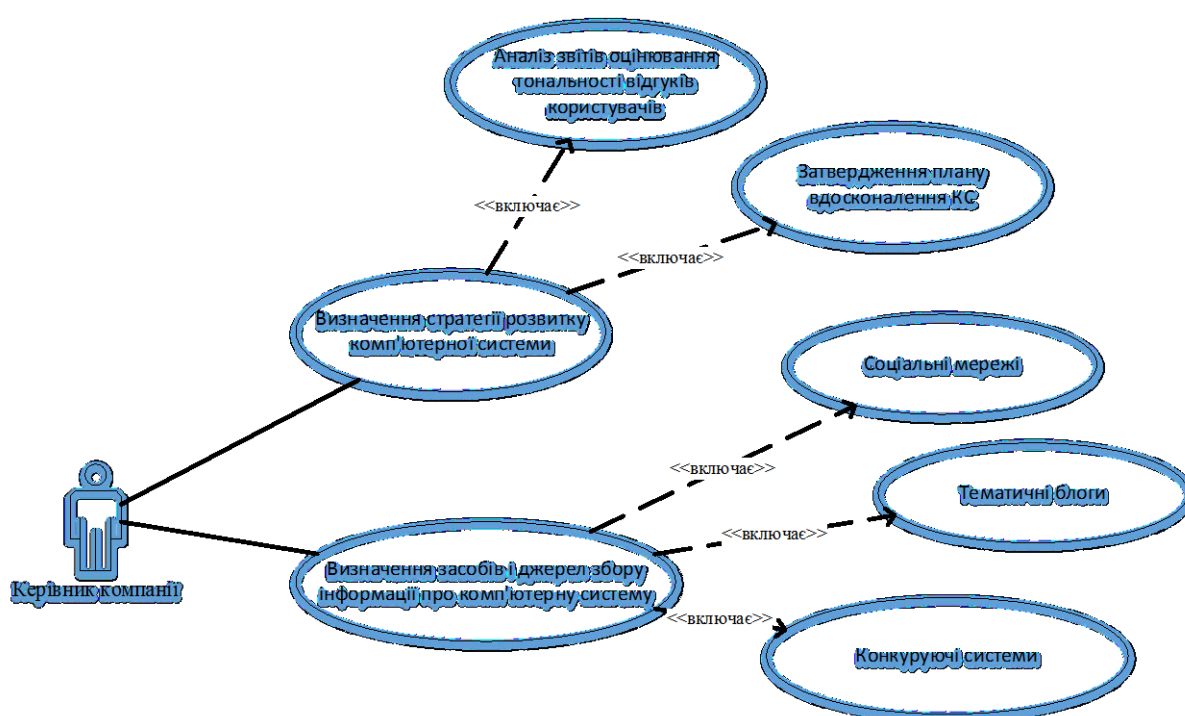


Рис. 3.3. Діаграма прецедентів «Керівник компанії»

Функції керівника компанії полягають у прийнятті рішень щодо подальшого розвитку комп'ютерної системи і включають наступні види діяльності:

- аналіз звітів щодо оцінювання тональності відгуків користувачів, які визначені аналітиками;
- затвердження плану розвитку комп'ютерної системи на основі звітів, наданих інженером з вдосконалення властивостей системи;
- визначення джерел збору даних на основі відгуків користувачів у соціальних мережах, тематичних блогах та на основі даних про конкуруючі комп'ютерні системи.

Визначивши основні функціональні вимоги та ролі користувачів системи виявлення та аналізу тональності текстової інформації, наступний крок полягає у проектуванні бази даних для зберігання даних, необхідних для прийняття рішення щодо вдосконалення комп'ютерних систем.

3.2. Проектування схеми бази даних при визначенні тональності текстової інформації

Перед тим, як проектувати базу даних, необхідно розробити шаблони та визначити тип інформації, яка необхідна для прийняття рішень щодо вдосконалення комп'ютерних систем. Оскільки, для визначення тональності текстової інформації запропоновано підхід, що базується на використанні аспектного семантичного аналізу, то структуру шаблону пропонується зобразити у вигляді, як показано на рис. 3.4.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Ідентифікатор	Посилання	Заголовок	Відгук	Дата	Час	Тип джерела	Тональність	Категорія	Мова	Країна	Регіон	Автор
2													
3													
4													
5													
6													
7													
8													
9													
10													
11													
12													
13													
14													

Рис. 3.4. Структура шаблону для аналізу тональності текстової інформації

Із структури шаблону видно, що обов'язковими полями в процесі аналізу та оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем є:

- ідентифікатор – унікальне значення за яким можна визначити конкретний відгук і кортеж взаємопов'язаних даних;
- посилання – містить інформацію про адресу джерела в мережі internet звідки одержано дані про комп'ютерну систему;
- заголовок – містить назву відгуку, зазвичай, береться перше речення з відгуку;
- відгук – власне тексту, що містить висловлювання про властивості або комп'ютерну систему в цілому;
- дата – дата написання відгуку;
- час – час написання відгуку;
- тип джерела – містить інформацію про тип джерела, наприклад, відео, зображення, текст та ін;
- тональність – класифікація відгуку, яка може приймати значення: позитивний, негативний, нейтральний;
- категорія – категорія аспекту, до якого належить відгук, наприклад, властивість або поведінка системи;

- мова – визначає якою мовою написано відгук;
- країна – дані про локацію, яку вказав користувач при реєстрації;
- регіон – деталізація інформації про джерело або відгук з географічної точки зору;
- автор – псевдонім користувача, який написав відгук.

Розробивши шаблон представлення відгуку і необхідний метаопис, можна спроектувати базу даних, для збереження та маніпулювання сутностями.

Сутності бази даних можна представити у вигляді реляційних відношень за забезпечити зв'язок між ними.

У табл. 3.1 наведено сутність `Computer_System` та її атрибути. Визначені поля інтерпретують наступним чином:

- `ID_Computer` – ідентифікатор комп'ютерної системи;
- `SystemName` – назва комп'ютерної системи;
- `AppArea` – сфера застосування комп'ютерної системи;
- `Description` – опис комп'ютерної системи

Таблиця 3.1

Computer_System

Назва поля	Тип поля	Примітка
<code>ID_Computer</code>	<code>int</code>	PK
<code>SystemName</code>	<code>varchar (200)</code>	
<code>AppArea</code>	<code>varchar (max)</code>	
<code>Description</code>	<code>varchar (max)</code>	

Для реалізації сутності представленої у вигляді табл. 3.1 написано SQL-запит, що приведений на рис. 3.5.

```
CREATE TABLE COMPUTER_SYSTEM  
(  
  ID_ComputerSystem int not null primary key identity (1,1),  
  SystemName varchar (200),  
  AppArea varchar (max),  
  [Description] varchar (max)  
)
```

Рис. 3.5. SQL-запит створення таблиці Computer_System

Оскільки, комп'ютерна система містить у собі програмні і апаратні компоненти, то необхідно визначити таблицю для збереження даних про складові (табл. 3.2).

Component

Назва поля	Тип поля	Примітка
ID_Component	int	PK
ID_ComputerSystem	int	FK
ComponentName	Varchar (max)	
Description	Varchar (max)	

У табл. 3.2 атрибути сутності Component представлено наступними полями:

- ID_Component – первинний ключ;
- ID_ComputerSystem – зовнішній ключ на таблицю Computer_System;
- ComponentName – назва компонента;
- Description – опис компонента.

Скрипт реалізації таблиці Component наведено на рис. 3.6.

```
CREATE TABLE Component
(
  ID_Component int not null primary key identity (1,1),
  ID_ComputerSystem int foreign key references Computer_System (ID_ComputerSystem),
  ComponentName varchar (max),
  [Description] varchar (max)
)
```

Рис. 3.6. Скрипт створення Component

Для маніпулювання властивостями комп'ютерної системи або її компонентів створено схему таблиці Features, яка показана у табл. 3.3.

Features

Назва поля	Тип поля	Примітка
ID_Feature	int	PK
ID_ComputerSystem	int	FK
ID_Component	int	FK
FeatureName	varchar (200)	
Description	varchar (max)	

Для реалізації табл. 3.3 засобами мови SQL написано код, що наведений на рис. 3.7.

```

CREATE TABLE Features
(
  ID_Feature int not null primary key identity (1,1),
  ID_ComputerSystem int foreign key references Computer_System (ID_ComputerSystem),
  ID_Component int foreign key references Component (ID_Component),
  FeatureName varchar (200),
  [Description] varchar (max)
)

```

Рис. 3.7. SQL-запит створення таблиці Features

Реляційне відношення Features містить зовнішні ключі на таблиці Computer_System та Component, а також поле назви властивості і її опис.

При визначенні тональності відгуків користувачів комп'ютерної системи, важливими є опис класів, які можуть приймати значення позитивності, негативності і нейтральності. Крім цього, для визначення емоційного забарвлення тексту, тональність може бути доповнена іншими класами. Для зберігання даних про класи тональності, визначено таблицю, що містить два поля:

- ID_Class – ідентифікатор класу;
- ClassName – назва класу.

На рис. 3.8 наведено програмний код створення цієї сутності.

```
CREATE TABLE SentimentClass
(
  ID_Class int not null primary key identity (1,1),
  ClassName varchar (50)
)
```

Рис. 3.8. SentimentClass

Для збереження інформації про джерела надходження відгуків, зокрема конкретного екземпляру соціальної мережі чи платформи перегляду відео, розроблено довідник, який представлено у вигляді табл. 3.4, а скрипт створення у середовищі SQL Server наведено на рис. 3.9.

Таблиця 3.4

Source

Назва поля	Тип поля	Примітка
ID_Source	int	PK
SourceName	varchar (200)	
Description	varchar (max)	

```
CREATE TABLE [Source]
(
  ID_Source int not null primary key identity (1,1),
  SourceName varchar (180),
  [Description] varchar (max)
)
```

Рис. 3.9. SQL-запит для створення таблиці Source

Для позначення мови, якою написано відгук користувача, створено довідник Language, що містить два поля:

- ID_Language – ідентифікатор мови;
- LanguageName – назва мови.

Створення таблиці Language у середовищі SQL Management Studio виконано шляхом написання запиту, наведеного на рис. 3.10.

```

CREATE TABLE [Language]
(
  ID_Language int not null primary key identity (1,1),
  LanguageName varchar (max),
.)

```

Рис. 3.10. SQL-запит створення таблиці Language

Користувач, який висловив думку відносно властивостей комп'ютерної системи, представлений сутністю User, опис якої наведено у табл. 3.5.

Таблиця 3.5

Users

Назва поля	Тип поля	Примітка
ID_User	int	PK
Login	varchar (200)	
ID_Source	int	FK
ID_Language	int	FK
Country	varchar (max)	
Region	varchar (max)	

Основні атрибути, якими описано користувача, призначені для його ідентифікації, визначення джерела одержання відгуку та мови, якої він написаний, приналежність користувача до країни та регіону.

На рис. 3.11 представлено фрагмент програмного коду для створення таблиці User.

```

CREATE TABLE [USER]
(
  ID_user int not null primary key identity (1,1),
  [Login] varchar (200),
  ID_Source int foreign key references [Source] (ID_Source),
  ID_Language int foreign key references [Language] (ID_Language),
  Country varchar (max),
  Region varchar (max)
)

```

Рис. 3.11. SQL-запит створення таблиці User

Для зберігання висловлювання користувача спроектовано таблицю Feedback, що наведена нижче (табл. 3.6)

Таблиця 3.6

Feedback

Назва поля	Тип поля	Примітка
ID_Feedback	int	PK
ID_Sentiment	int	FK
ID_User	int	FK
FeedbackText	varchar(max)	
FeedbackDate	date	
FeedbackTime	time	

У таблиці Feedback наведено наступні поля:

- ID_Feedback – ідентифікатор відгуку про комп'ютерну систему;
- ID_Sentiment – зовнішній ключ для встановлення тональності відгуку;
- ID_User – зовнішній ключ, що ідентифікує користувача;
- FeedbackText – посилання або текст відгуку користувача;
- FeedbackDate – дата написання висловлювання користувачем;
- FeedbackTime – час створення відгуку.

SQL- запит для створення таблиці Feedback наведено на рис. 3.12.

```
CREATE TABLE Feedback
(
  ID_Feedback int not null primary key identity (1,1),
  ID_Sentiment int not null foreign key references SentimentClass (ID_Class),
  ID_User int not null foreign key references [User] (ID_User),
  FeedbackText varchar (max),
  FeedbackDate date,
  FeedbackTime time,
)
```

Рис. 3.12. Feedback

При аналізі відгуку користувача про комп'ютерну систему або її компонент, необхідно визначити сутність або сутності про які іде мова. Таблицю Сутність описано так, як наведено у табл. 3.7.

Таблиця 3.7

Entity

Назва поля	Тип поля	Примітка
ID_Entity	int	PK
ID_Feedback	int	FK
EntityName	varchar(200)	

Визначена у відгуку сутність пов'язана зовнішнім ключем ID_Feedback з таблицею Feedback. Для реалізації таблиці Сутність написано запит, який показано на рис. 3.13

```
CREATE TABLE Entity
(
  ID_Entity int not null primary key identity (1,1),
  EntityName varchar (200),
  ID_Feedback int foreign key references Feedback (ID_Feedback),
```

Рис. 3.13. Entity

Для виявлення аспектів сутності про яку іде мова у висловлюванні побудовано табл. 3.8.

Таблиця 3.8

Aspect

Назва поля	Тип поля	Примітка
ID_Aspect	int	PK
ID_Class	int	FK
AspectName	varchar(200)	
Description	varchar(max)	

У табл. 3.8 полями виступають:

- ID_Aspect – ідентифікатор аспекту;
- ID_Class – значення тональності аспекту (зовнішній ключ на таблиці SentimentClass);
- AspectName – назва аспекту;
- Description – додатковий опис виявленого аспекту.

На рис. 3.14 наведено запит для створення табл. 3.8.

```
CREATE TABLE Aspect
(
  ID_Aspect int not null primary key identity (1,1),
  ID_Class int not null foreign key references SentimentClass (ID_Class),
  AspectName varchar (200),
  [Description] varchar (max),
)
```

Рис. 3.14. Aspect

Для забезпечення зв'язку між властивостями і аспектами створено проміжну табл. 3.9.

Таблиця 3.9

Feature_Aspect

Назва поля	Тип поля	Примітка
ID_Feature_Aspect	int	PK
ID_Feature	int	FK
ID_Aspect	int	FK

Реалізація запиту на створення Feature_Aspect наведено на рис. 3.15.

```
create table Feature_Aspect
(
  ID_Feature_Aspect int not null primary key identity (1,1),
  ID_Feature int foreign key references Features (ID_Feature),
  ID_Aspect int foreign key references Aspect (ID_Aspect)
)
```

Рис. 3.15. Створення таблиці Feature_Aspect

Оскільки, існує імовірність того, що один і той же аспект може описувати кілька сутностей, тому варто створити проміжну таблицю Entity_Aspect, структура якої наведена у табл. 3.10 та на рис. 3.16.

Таблиця 3.10

Entity_Aspect

Назва поля	Тип поля	Примітка
ID_Entity_Aspect	int	PK
ID_Entity	int	FK
ID_Aspect	int	FK

```
CREATE TABLE Entity_Aspect
(
  ID_Entity_Aspect int not null primary key identity (1,1),
  ID_Entity int foreign key references Entity (ID_Entity),
  ID_Aspect int foreign key references Aspect (ID_Aspect)
)
```

Рис. 3.16. Проміжна таблиця Entity_Aspect

Оскільки, запропонованим методом визначено необхідність використання словників слів для оцінювання тональності аспектів, то його представлено у вигляді табл. 3.11., а SQL-запит її створення на рис. 3.17.

Таблиця 3.11

EvalWord

Назва поля	Тип поля	Примітка
ID_EvalWord	int	PK
ID_Class	int	FK
Word	varchar(100)	


```

CREATE TABLE EvalWord
(
ID_EvalWord int not null primary key identity (1,1),
ID_Class int foreign key references SentimentClass (ID_Class),
Word varchar (100),
)

```

Рис. 3.17. Програмний код створення таблиці EvalWord

Для встановлення відповідності та зв'язку між словами, які мітять оцінку аспекту і самого аспекту спроектовано таблицю 3.12.

Таблиця 3.12

Aspect_EvalWord

Назва поля	Тип поля	Примітка
ID_Aspect_Word	int	PK
ID_EvalWord	int	FK
ID_Aspect	int	FK

На рис. 3.18 наведено скрипт генерації табл. 3.12.

```

CREATE TABLE Aspect_EvalWords
(
ID_Aspect_Word int not null primary key identity (1,1),
ID_Aspect int foreign key references Aspect (ID_Aspect),
ID_EvalWord int foreign key references EvalWord (ID_EvalWord)
)

```

Рис. 3.18. Скрипт Aspect_EvalWord

Ще одна проміжна таблиця User_Source встановлює зв'язок між користувачем і його профілем у соціальній мережі, або іншому джерелі. SQL-запит створення таблиці User_Source наведено на рис. 3.19, сама таблиця представлена у табл. 3.13.

```

create table User_Source
(
ID_User_Source int not null primary key identity (1,1),
ID_User int foreign key references [User] (ID_User),
ID_Source int foreign key references [Source] (ID_Source)
)

```

Рис. 3.19. Таблиця User_Source

Таблиця 3.13

User_Source

Назва поля	Тип поля	Примітка
ID_User_Source	int	PK
ID_User	int	FK
ID_Source	int	FK

Після створення усіх необхідних таблиць спроектовано ER-діаграму, що наведена на рис. 3.20. Дана діаграма показує визначені у предметній області сутності та зв'язки між ними.

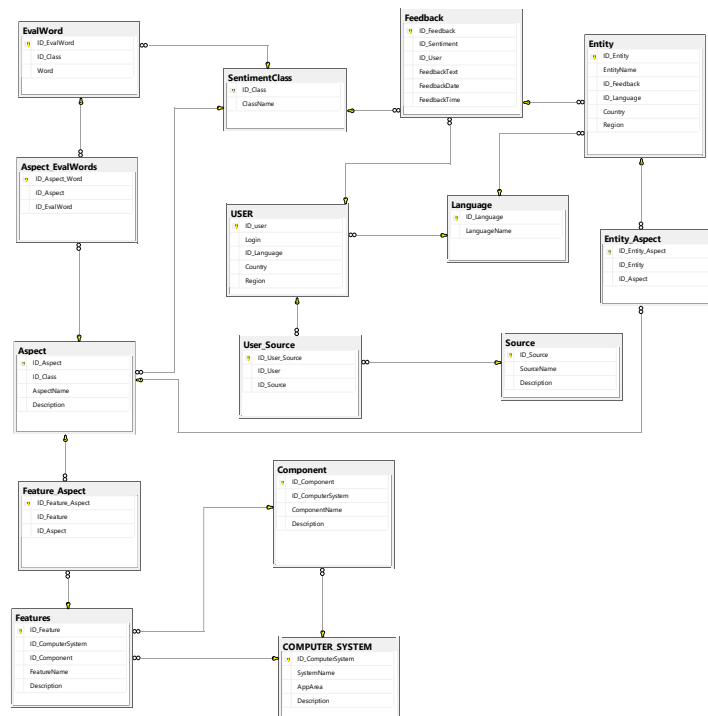


Рис. 3.20. Схема бази даних для управління процесом виявлення та аналізу тональності текстової інформації при вдосконаленні комп'ютерних систем

Застосування спроектованої бази даних забезпечує відповідність вимогам щодо нормалізації і надлишковості реляційних відношень, дозволяє гнучко маніпулювати даними і процес виявлення та оцінювання тональності відгуків про якість комп'ютерної системи та її компонентів.

3.3. Проектування архітектури системи виявлення та оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем

Для представлення архітектури системи виявлення та оцінювання тональності відгуків користувачів про комп'ютерні системи запропоновано скористатись шарами Фаулера. У загальному випадку архітектуру розроблюваної системи можна зобразити як показано на рис. 3.21.

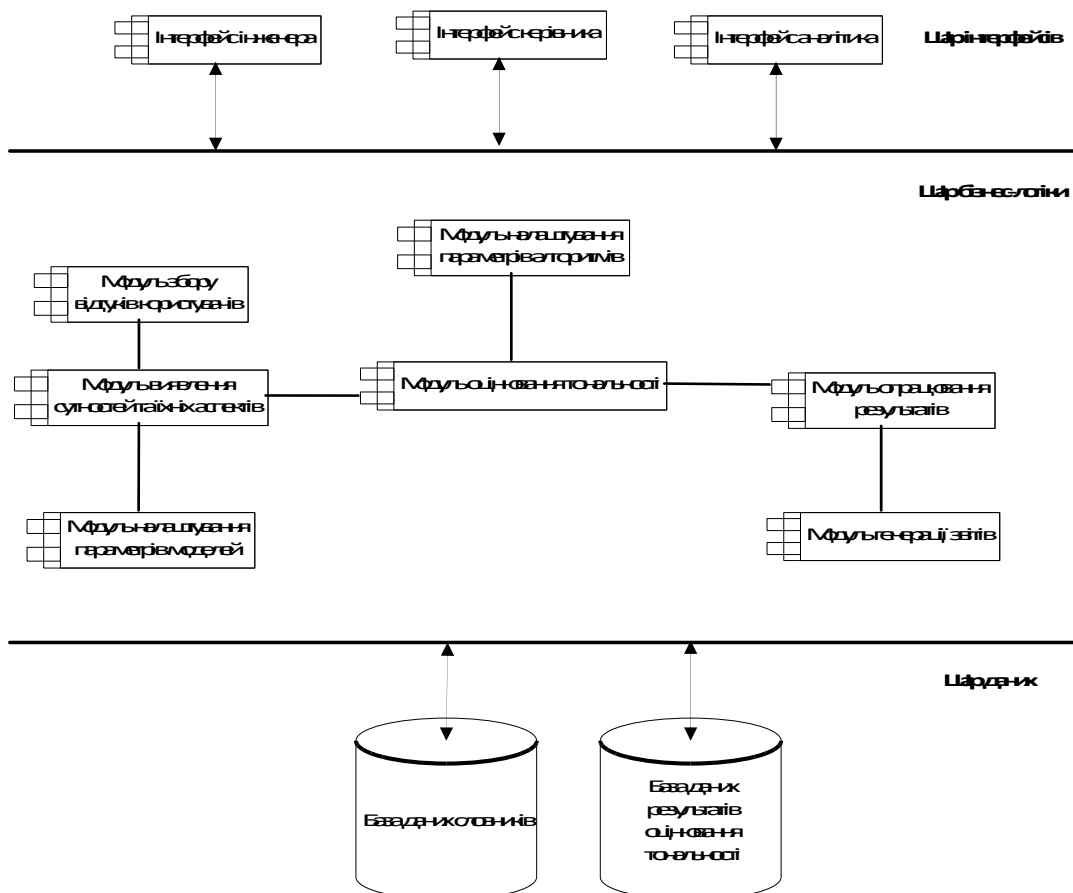


Рис. 3.21. Архітектура системи управління процесами виявлення та оцінювання тональності текстової інформації

На рівні шару користувацьких інтерфейсів визначено три інтерфейси – інтерфейс інженера, інтерфейс аналітика та інтерфейс керівника компанії.

Інтерфейс інженера з вдосконалення комп'ютерних систем повинен реалізовувати функції, які наведено на рис. 3.1 і мати відповідні елементи керування для виконання функціональних обов'язків інженера.

Інтерфейс аналітика повинен бути адаптованим та відповідати за виконання функціональних можливостей, які наведені на рис. 3.2., а інтерфейс керівника компанії – орієнтований на забезпечення виконання функціональних вимог, представлених на рис. 3.3.

У шарі бізнес-логіки передбачено модуль для збору відгуків користувачів про комп'ютерну систему або її компоненти. Даний модуль можна використовувати як результат розробки сторонньої компанії шляхом доступу до сервісу через API.

Модуль виявлення сутностей розміщується на стороні cloud сервісу і передбачає практичне використання відповідних моделей виявлення сутностей, зокрема TF-IDF. Оскільки, для виявлення та аналізу сутностей може використовуватись не одна модель, то забезпечено доступ до модуля налаштування параметрів моделей. Це дає змогу підвищити повноту у процесів добування сутностей та їх властивостей.

Модуль оцінювання тональності висловлювань забезпечує організацію процесу розпізнавання позитивних, негативних і нейтральних аспектів з можливістю налаштування параметрів алгоритмів оцінювання та методів класифікації.

Модуль опрацювання результатів дає змогу оцінювати повноту і достовірність одержаних оцінок як автоматичним шляхом, так і за допомогою експертного оцінювання.

Модуль генерації звітів дозволяє згенерувати документацію щодо результатів оцінювання тональності текстової інформації при вдосконаленні властивосте комп'ютерних систем.

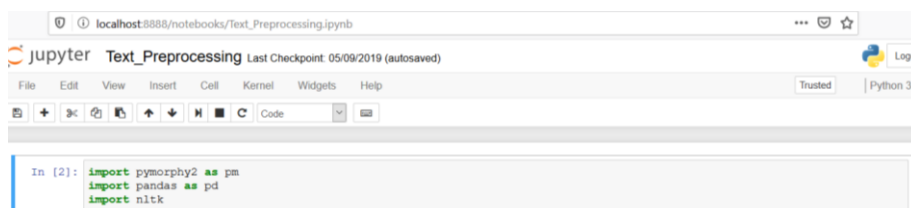
База даних словників містить слова, за допомогою яких проводиться класифікація аспектів за класами позитивний, негативний, нейтральний.

База даних результатів оцінювання тональності містить рекомендації щодо вдосконалення властивостей комп'ютерних систем.

Як приклад, наведемо фрагмент реалізації системи оцінювання тональності текстової інформації щодо попередньої синтаксичної обробки кириличних відгуків. Для скористаємось середовищем Jupyter Notebook, мовою Python і кількома бібліотеками у сфері data science, зокрема

- Rymorphy2 – бібліотека синтаксичного розбору речень, написаних кирилицею;
- Pandas – бібліотека, яка в даному випадку використовується для представлення структур даних;
- Nltk – бібліотека, що в даному випадку використовується для визначення міри подібності слів із застосуванням embedding.

На рис. 3.22 наведено підключення наведених вище бібліотек.



```
In (2): import pymorphy2 as pm
import pandas as pd
import nltk
```

Рис. 3.22. Підключення бібліотек

Підключивши необхідні бібліотеки, необхідно провести зчитування відгуків про комп'ютерні системи із файлу з розширенням .csv і зберегти їх у фрейм даних. На рис. 3.23 наведено функцію для зчитування даних з файлу Test.csv.

```
#load posts from dataset
def read_csv(string):
    df=pd.read_csv(string, sep=';')
    return df
df=read_csv('Test.csv')
print(df.head(5))
```

Рис. 3.23. Зчитування даних з файлу Test.csv

У результаті зчитування дані поміщаються у фрейм даних (рис. 3.24)

	ID	Текст
0	63887434919	+shkoolley _ вот про кастомизацию грамотно ска...
1	63886207026	Самсунгёб со стажем репортинг. Не покупайте но...
2	63886701348	Батарей гавно. (владею с8 больше года)
3	63887434787	Тут разница в возрасте ещё играет и чем челове...
4	63891859770	Gear IconX 2018 поолное гавно!Я это говорю, п...

Рис. 3.24. Дані з фрейму

Для того, щоб визначати сутності та аспекти необхідно привести усі слова у реченнях до нижнього регістру. Функція, яка реалізує таке приведення наведена на рис. 3.25.

```
#make text lower
def make_lower(df):
    for i in range(0,df.shape[0]):
        df.iloc[i,1]=df.iloc[i,1].lower()
    return df
make_lower(df)
print (df['Текст'].head(5))
```

Рис. 3.25. Функція приведення речень до нижнього регістру

Наступний крок полягає у токенизації тексту, тобто визначенні окремих слів і їх властивостей (рис. 3.26).

```
In [98]: #create tokenized list of posts
posts_list=[]
for i in range(0,df.shape[0]):
    posts_list.append(df.iloc[i,1])
word_tokens=[]
for j in range(0,len(posts_list)):
    word_tokens.append(nltk.word_tokenize(posts_list[j],language="russian"))
#display tokenized posts
for i in range (0,len(word_tokens)):
    print(word_tokens[i])
```

```
[ '+shkoolley', '_', 'вот', 'про', 'кастомизацию', 'грамотно', 'сказал', '!', '!', '!', 'кто', 'то', 'собой', 'занимается',
',', '!', 'а', 'кто', 'сидит', 'на', 'форумах', 'и', 'кастомные', 'прошивки', 'ишет', '!', '!', 'тема', 'забавная', 'была',
',', 'лет', '13-14', 'назад', '!', '!', 'сиенс', 'был', 'сл75', '!', '!', 'разблокирован', '!', '!', 'можно', 'было', 'иконки', 'по',
'менять', 'прям', 'в', 'мобиле', '!', '!', 'каких', 'нибудь', 'плхх', 'накатить', 'без', 'компа', '!', '!', 'рабочий', 'стол',
',', 'от', 'винды', 'с', 'курсором', '!', '!', 'через', 'браузер', '!', '!', 'но', 'функционирующий', '!', '!', 'здорово', 'было', '!',
',', 'но', 'было', 'это', 'давно', '!', '!', 'сейчас', 'же', 'этим', 'заниматься', 'лень', '!', '!', 'да', 'и', 'смысла', 'не', 'в',
'ижу', '!', '!', 'на', 'айфоне', 'все', 'красиво', 'из', 'коробки', '!', '!', 'работает', 'быстро', '!', '!', 'плюс', 'дизайн', 'и',
',', 'материалы', '!', '!', 'всегда', 'топил', 'за', 'дизайн', '!', '!', 'чек', 'за', 'функционал', '!', '!', 'месенджеры', '!', '!', 'му',
'зыка', '!', '!', 'именно', 'звук', '!', '!', 'лучше', 'чек', 'у', 'большинства', '!', '!', 'книги', '!', '!', 'аудиокниги', '!', '!',
',', 'дисплей', 'и', 'хорошая', 'камера', '!', '!', '!', 'под', 'мои', 'задачи', '!', '!', '!', 'все', 'это', 'есть', 'из', 'коробк',
'и', '!', '!', 'есть', 'минус', '!', '!', 'будильник', 'кривой', '!', '!', 'вот', 'бесит', '!', '!', 'и', 'все', '!', '!', 'а', 'теперь', 'че',
'го', 'не', 'хватает', '!', '!', 'стилус', 'как', 'в', 'самсунге', 'ноут', '!', '!', 'как', 'вторую', 'мобилу', 'буду', 'брать',
',', 'ноут9', '!', '!', 'наличие', 'двух', 'сим', 'карт', 'не', 'интересно', '!', '!', 'это', 'на', 'всякий', 'случай', '!',
',', '!', 'основной', 'номер', 'в', 'отдельной', 'трубке', '!', '!', 'так', 'нужно', '!', '!', 'сушь', 'именно', 'в', 'стилусе',
',', '!', 'очень', 'много', 'заметок', 'пишу', '!', '!', 'печатать', 'долго', '!', '!', 'плюс', 'иногда', 'нужно', 'зарисовать',
',', '!', 'и', 'вот', 'тут', 'ноут', 'вне', 'конкуренции', '!', '!', 'действительно', 'не', 'мобила', '!', '!', 'а', 'рабочая', 'з',
'верюга', '!', '!', 'андроид', 'не', 'плохая', 'операционка', '!', '!', 'были', 'бы', 'все', 'иконки', 'в', 'одном', 'стиле',
',', 'и', 'поменьше', 'настроек', '!', '!', 'то', 'может', 'и', 'перешёл', 'бы', 'обратно', '!', '!', 'а', 'так', '!', '!', 'слишком',
',', 'долго', 'трубку', 'копьять', 'нужно', '!', '!', 'наверное', 'так', 'скажу', '!', '!', 'андроид', 'возможно', 'лучшая', 'опе
```

Рис. 3.26. Функція та результати токенизації тексту

Для побудови правил та навчання нейромережі необхідно провести визначення частин мови у токенозованому тексті. Тіло функції визначення частин мови наведено на рис. 3.27.

```
#create list of Part-of Speech for all posts
morph = pm.MorphAnalyzer()
vec_pos=[]
vec=""
print(len(word_tokens))
for i in range(0,1):
  for j in range(0,len(word_tokens[i])):
    vec=vec+', '+str(word_tokens[i][j])+', '+str(morph.parse(word_tokens[i][j])[0].tag.POS)+' '
    vec_pos.append(vec)
  print(vec)

for i in range(0,len(vec_pos)):
  print(vec_pos[i])
```

```
12043
, {+shkoolley, None}, {, None}, {vor, PRCL}, {про, PREP}, {кастомизацию, NOUN}, {грамотно, ADVB}, {сказал, VERB}, {!, None}, {, None}, {кто, NPRO}, {то, CONJ}, {свой, NPRO}, {занимается, VERB}, {, None}, {а, CONJ}, {кто, NPRO}, {сидит, VERB}, {на, PREP}, {форумах, NOUN}, {и, CONJ}, {кастомия, ADJF}, {прошивки, NOUN}, {имет, VERB}, {, None}, {тема, NOUN}, {забавная, ADJF}, {была, VERB}, {лет, NOUN}, {13-14, None}, {и, AZAD}, {авад, ADVB}, {, None}, {сименс, NOUN}, {был, VERB}, {сл75, None}, {, None}, {разблокирован, PRS}, {, None}, {можно, PRED}, {было, VERB}, {и, KONKI}, {конки, NOUN}, {поменять, INFN}, {пряч, PRCL}, {в, PREP}, {мобиле, NOUN}, {, None}, {каких, ADJF}, {нибудь, VERB}, {плюх, INTJ}, {накатить, IN FN}, {без, PREP}, {компа, NOUN}, {, None}, {рабочий, NOUN}, {стол, NOUN}, {от, PREP}, {винды, NOUN}, {с, PREP}, {курсором, NOUN}, {, None}, {ч ерез, PREP}, {браузер, NOUN}, {, None}, {но, CONJ}, {функционалукий, PRTF}, {, None}, {здорово, INTJ}, {было, VERB}, {!, None}, {но, CON J}, {было, VERB}, {это, PRCL}, {давно, ADVB}, {!, None}, {сейчас, ADVB}, {же, PRCL}, {этими, NPRO}, {заниматься, INFN}, {лень, NOUN}, {, Non e}, {да, PRCL}, {и, CONJ}, {смысла, NOUN}, {не, PRCL}, {вижу, VERB}, {!, None}, {на, PREP}, {айфоне, NOUN}, {все, ADJF}, {красиво, ADVB}, {из, PR EP}, {коробки, NOUN}, {!, None}, {работает, VERB}, {быстро, ADVB}, {, None}, {плюс, CONJ}, {дизайн, NOUN}, {и, CONJ}, {материалы, NOUN}, {!, N one}, {всегда, ADVB}, {топил, VERB}, {за, PREP}, {дизайн, NOUN}, {, None}, {чем, CONJ}, {за, PREP}, {функционал, NOUN}, {!, None}, {месенджер ы, NOUN}, {, None}, {музыка, NOUN}, {, None}, {именно, PRCL}, {звук, NOUN}, {, None}, {лучше, COMP}, {чем, CONJ}, {у, PREP}, {большинства, NO UN}, {, None}, {, None}, {книги, NOUN}, {, None}, {аудиокниги, NOUN}, {, None}, {дисплей, NOUN}, {и, CONJ}, {хорошая, ADJF}, {камера, NOU N}, {!, None}, {, None}, {под, PREP}, {мои, ADJF}, {задачи, NOUN}, {, None}, {, None}, {все, ADJF}, {это, PRCL}, {есть, INTJ}, {из, PREP}, {кор обки, NOUN}, {!, None}, {, None}, {есть, INTJ}, {изнус, CONJ}, {, None}, {будильник, NOUN}, {кривой, NOUN}, {!, None}, {вот, PRCL}, {бесит, VERB}, {!, No ne}, {и, CONJ}, {все, ADJF}, {, None}, {а, CONJ}, {теперь, ADVB}, {чего, NPRO}, {не, PRCL}, {хватает, VERB}, {!, None}, {стилус, NOUN}, {как, CO NJ}, {в, PREP}, {самсунге, NOUN}, {ноут, NOUN}, {!, None}, {как, CONJ}, {вторую, ADJF}, {мобилу, NOUN}, {буду, VERB}, {буду9, No ne}, {!, None}, {наличие, NOUN}, {двух, NUMR}, {сии, NOUN}, {карт, NOUN}, {не, PRCL}, {интересно, ADVB}, {!, None}, {это, PRCL}, {на, PREP}, {в сякий, ADJF}, {случай, NOUN}, {, None}, {, None}, {основной, ADJF}, {номер, NOUN}, {в, PREP}, {отдельной, ADJF}, {трубка, NOUN}, {!, None}, {в ак, CONJ}, {нужно, PRED}, {!, None}, {суть, VERB}, {именно, PRCL}, {з, PREP}, {стилусе, NOUN}, {!, None}, {очень, ADVB}, {много, ADVB}, {замено к, NOUN}, {пешу, VERB}, {, None}, {печатать, INFN}, {долго, ADVB}, {, None}, {плюс, CONJ}, {иногда, ADVB}, {нужно, PRED}, {зарисовать, INF N}, {!, None}, {и, CONJ}, {вот, PRCL}, {тут, ADVB}, {ноут, NOUN}, {эне, PREP}, {конкуренции, NOUN}, {, None}, {действительно, ADVB}, {не, PR CL}, {мобила, NOUN}, {, None}, {а, CONJ}, {рабочая, ADJF}, {зверга, NOUN}, {!, None}, {андроид, NOUN}, {не, PRCL}, {плохая, ADJF}, {операцион ка, NOUN}, {, None}, {были, VERB}, {бы, PRCL}, {все, ADJF}, {иконки, NOUN}, {в, PREP}, {одном, ADJF}, {стиле, NOUN}, {и, CONJ}, {поменьше, COM P}, {настроек, NOUN}, {, None}, {то, CONJ}, {может, VERB}, {и, CONJ}, {перешел, VERB}, {бы, PRCL}, {обратно, ADVB}, {, None}, {а, CONJ}, {так, CONJ}, {, None}, {олишком, ADVB}, {долго, ADVB}, {трубку, NOUN}, {жоврят, INFN}, {нужно, PRED}, {, None}, {насерное, ADVB}, {так, CON J}, {скажу, VERB}, {, None}, {андроид, NOUN}, {возможно, CONJ}, {лучшая, ADJF}, {операционка, NOUN}, {, None}, {при, PREP}, {условии, NOU N}, {наличие, NOUN}, {времени, NOUN}, {и, CONJ}, {желания, NOUN}, {её, ADJF}, {жоврят, INFN}, {, None}, {если, CONJ}, {желания, NOUN}, {и, C ONJ}, {времени, NOUN}, {нет, PRED}, {, None}, {то, CONJ}, {лучше, COMP}, {спл, None}, {, None}
, {+shkoolley, None}, {, None}, {vor, PRCL}, {про, PREP}, {кастомизацию, NOUN}, {грамотно, ADVB}, {сказал, VERB}, {!, None}, {, None}, {кто, NPRO}, {то, CONJ}, {свой, NPRO}, {занимается, VERB}, {, None}, {а, CONJ}, {кто, NPRO}, {сидит, VERB}, {на, PREP}, {форумах, NOUN}, {и, CONJ}, {кастомия, ADJF}, {прошивки, NOUN}, {имет, VERB}, {, None}, {тема, NOUN}, {забавная, ADJF}, {была, VERB}, {лет, NOUN}, {13-14, None}, {и, AZAD}, {авад, ADVB}, {, None}, {сименс, NOUN}, {был, VERB}, {сл75, None}, {, None}, {разблокирован, PRS}, {, None}, {можно, PRED}, {было, VERB}, {и, KONKI}, {конки, NOUN}, {поменять, INFN}, {пряч, PRCL}, {в, PREP}, {мобиле, NOUN}, {, None}, {каких, ADJF}, {нибудь, VERB}, {плюх, INTJ}, {накатить, IN FN}, {без, PREP}, {компа, NOUN}, {, None}, {рабочий, NOUN}, {стол, NOUN}, {от, PREP}, {винды, NOUN}, {с, PREP}, {курсором, NOUN}, {, None}, {ч ерез, PREP}, {браузер, NOUN}, {, None}, {но, CONJ}, {функционалукий, PRTF}, {, None}, {здорово, INTJ}, {было, VERB}, {!, None}, {но, CON J}, {было, VERB}, {это, PRCL}, {давно, ADVB}, {!, None}, {сейчас, ADVB}, {же, PRCL}, {этими, NPRO}, {заниматься, INFN}, {лень, NOUN}, {, Non e}, {да, PRCL}, {и, CONJ}, {смысла, NOUN}, {не, PRCL}, {вижу, VERB}, {!, None}, {на, PREP}, {айфоне, NOUN}, {все, ADJF}, {красиво, ADVB}, {из, PR EP}, {коробки, NOUN}, {!, None}, {работает, VERB}, {быстро, ADVB}, {, None}, {плюс, CONJ}, {дизайн, NOUN}, {и, CONJ}, {материалы, NOUN}, {!, N one}, {всегда, ADVB}, {топил, VERB}, {за, PREP}, {дизайн, NOUN}, {, None}, {чем, CONJ}, {за, PREP}, {функционал, NOUN}, {!, None}, {месенджер ы, NOUN}, {, None}, {музыка, NOUN}, {, None}, {именно, PRCL}, {звук, NOUN}, {, None}, {лучше, COMP}, {чем, CONJ}, {у, PREP}, {большинства, NO UN}, {, None}, {, None}, {книги, NOUN}, {, None}, {аудиокниги, NOUN}, {, None}, {дисплей, NOUN}, {и, CONJ}, {хорошая, ADJF}, {камера, NOU N}, {!, None}, {, None}, {под, PREP}, {мои, ADJF}, {задачи, NOUN}, {, None}, {, None}, {все, ADJF}, {это, PRCL}, {есть, INTJ}, {из, PREP}, {кор обки, NOUN}, {!, None}, {, None}, {есть, INTJ}, {изнус, CONJ}, {, None}, {будильник, NOUN}, {кривой, NOUN}, {!, None}, {вот, PRCL}, {бесит, VERB}, {!, No ne}, {и, CONJ}, {все, ADJF}, {, None}, {а, CONJ}, {теперь, ADVB}, {чего, NPRO}, {не, PRCL}, {хватает, VERB}, {!, None}, {стилус, NOUN}, {как, CO NJ}, {в, PREP}, {самсунге, NOUN}, {ноут, NOUN}, {!, None}, {как, CONJ}, {вторую, ADJF}, {мобилу, NOUN}, {буду, VERB}, {буду9, No ne}, {!, None}, {наличие, NOUN}, {двух, NUMR}, {сии, NOUN}, {карт, NOUN}, {не, PRCL}, {интересно, ADVB}, {!, None}, {это, PRCL}, {на, PREP}, {в сякий, ADJF}, {случай, NOUN}, {, None}, {, None}, {основной, ADJF}, {номер, NOUN}, {в, PREP}, {отдельной, ADJF}, {трубка, NOUN}, {!, None}, {в ак, CONJ}, {нужно, PRED}, {!, None}, {суть, VERB}, {именно, PRCL}, {з, PREP}, {стилусе, NOUN}, {!, None}, {очень, ADVB}, {много, ADVB}, {замено к, NOUN}, {пешу, VERB}, {, None}, {печатать, INFN}, {долго, ADVB}, {, None}, {плюс, CONJ}, {иногда, ADVB}, {нужно, PRED}, {зарисовать, INF N}, {!, None}, {и, CONJ}, {вот, PRCL}, {тут, ADVB}, {ноут, NOUN}, {эне, PREP}, {конкуренции, NOUN}, {, None}, {действительно, ADVB}, {не, PR CL}, {мобила, NOUN}, {, None}, {а, CONJ}, {рабочая, ADJF}, {зверга, NOUN}, {!, None}, {андроид, NOUN}, {не, PRCL}, {плохая, ADJF}, {операцион ка, NOUN}, {, None}, {были, VERB}, {бы, PRCL}, {все, ADJF}, {иконки, NOUN}, {в, PREP}, {одном, ADJF}, {стиле, NOUN}, {и, CONJ}, {поменьше, COM P}, {настроек, NOUN}, {, None}, {то, CONJ}, {может, VERB}, {и, CONJ}, {перешел, VERB}, {бы, PRCL}, {обратно, ADVB}, {, None}, {а, CONJ}, {так, CONJ}, {, None}, {олишком, ADVB}, {долго, ADVB}, {трубку, NOUN}, {жоврят, INFN}, {нужно, PRED}, {, None}, {насерное, ADVB}, {так, CON J}, {скажу, VERB}, {, None}, {андроид, NOUN}, {возможно, CONJ}, {лучшая, ADJF}, {операционка, NOUN}, {, None}, {при, PREP}, {условии, NOU N}, {наличие, NOUN}, {времени, NOUN}, {и, CONJ}, {желания, NOUN}, {её, ADJF}, {жоврят, INFN}, {, None}, {если, CONJ}, {желания, NOUN}, {и, C ONJ}, {времени, NOUN}, {нет, PRED}, {, None}, {то, CONJ}, {лучше, COMP}, {спл, None}, {, None}
```

Рис. 3.27. Визначення частин мови для слів

Далі потрібно провести embedding слів, для цього їх необхідно записати в базу даних і кількісно виразити компоненти векторів (рис. 3.28).

```
In [92]: word_tokens
vec_parse=[]
vec_parse_str=""
for i in range(0,1):
  for j in range(0,len(word_tokens[i])):
    vec_parse_str=vec_parse_str+', '+str(word_tokens[i][j])+', '+str(morph.parse(word_tokens[i][j]))
  vec_parse.append(vec_parse_str)
  print(vec_parse)
```

```
[" (+shkoolley, [Parse(word='+shkoolley', tag=OpenCorporaTag('LATN'), normal_form='+shkoolley', score=1.0, methods_stack=(<LatinAnalyzer>, '+shkoolley'))], [_, [Parse(word='_', tag=OpenCorporaTag('FNCT'), normal_form=' ', score=1.0, methods_stack=(<PunctuationAnalyzer>, ' '))], [vor, [Parse(word='vor', tag=OpenCorporaTag('PRCL'), normal_form='vor', score=1.0, methods_stack=(<DictionaryAnalyzer>, 'vor', 22, 0))], [про, [Parse(word='про', tag=OpenCorporaTag('PREP'), normal_form='про', score=1.0, methods_stack=(<DictionaryAnalyzer>, 'про', 24, 0))], [кастомизацию, [Parse(word='кастомизация', tag=OpenCorporaTag('NOUN_inan_femn_sing_accs'), normal_form='кастомизация', score=1.0, methods_stack=(<FakeDictionaryAnalyzer>, 'кастомизация', 40, 3), (<KnownSuffixAnalyzer>, 'зация')], [грамотно, [Parse(word='грамотно', tag=OpenCorporaTag('ADVB'), normal_form='грамотно', score=0.5, methods_stack=(<DictionaryAnalyzer>, 'грамотно', 3, 0))], [Parse(word='грамотно', tag=OpenCorporaTag('ADJF_qual_neut_sing'), normal_form='грамотный', score=0.5, methods_stack=(<DictionaryAnalyzer>, 'грамотно', 86, 56))], [сказал, [Parse(word='сказал', tag=OpenCorporaTag('VERB_ref_tran_mas_sing_past_ind'), normal_form='сказал', score=1.0, methods_stack=(<DictionaryAnalyzer>, 'сказал', 638, 1))], [!, [Parse(word='!', tag=OpenCorporaTag('FNCT'), normal_form='!', score=1.0, methods_stack=(<PunctuationAnalyzer>, '!'))], [], [Parse(word=''), tag=OpenCorporaTag('FNCT'), normal_form='', score=1.0, methods_stack=(<PunctuationAnalyzer>, ' '))], [кто, [Parse(word='кто', tag=OpenCorporaTag('NPRO_mas_sing_nonn'), normal_form='кто', score=1.0, methods_stack=(<DictionaryAnalyzer>, 'кто', 174, 2, 0))], [то, [Parse(word='то', tag=OpenCorporaTag('CONJ'), normal_form='то', score=0.5, methods_stack=(<DictionaryAnalyzer>, 'то', 20, 0))], [Parse(word='то', tag=OpenCorporaTag('ADJF_Apro_Subx_Amph_neut_sing_accs'), normal_form='тор', score=0.3, methods_stack=(<DictionaryAnalyzer>, 'то', 2900, 16))], [Parse(word='то', tag=OpenCorporaTag('PRCL'), normal_form='то', score=0.1, methods_stack=(<DictionaryAnalyzer>, 'то', 22, 0))], [Parse(word='то', tag=OpenCorporaTag('ADJF_Apro_Subx_Amph_neut_sing_nonn'), normal_form='tor', score=0.1, methods_stack=(<DictionaryAnalyzer>, 'то', 2900, 13))], [
```

Рис. 3.28. Embedding

У даній частині роботи наведено основні функції препроцесингу текстових даних і загальну організацію системи виявлення та аналізу текстової інформації при вдосконаленні комп'ютерних систем. Це дає змогу реалізувати повноцінний комерційний продукт для виявлення і аналізу тональності тексту. Наведемо основні результати роботи розробленої системи.

Результати апробації запропонованого методу, моделі і системи наведено у додатку Б.

3.4. Висновки до розділу

1. При розробці системи виявлення та аналізу тональності текстової інформації для вдосконаленні комп'ютерних систем побудовано діаграми варіантів використання, що дало змогу визначити функціональні вимоги до системи та ролі користувачів.

2. Розроблено схему бази даних для зберігання та маніпулювання даними щодо оцінювання тональності текстової інформації (відгуків користувачів), що дало змогу забезпечити гнучкість та масштабованість процесу вдосконалення комп'ютерних систем.

3. Спроектовано архітектуру системи виявлення та оцінювання тональності відгуків користувачів про комп'ютерну систему або її компоненти, що дало змогу реалізувати програмні модулі керування параметрами алгоритмів і моделей машинного навчання.

РОЗДІЛ 4 ОБГРУНТУВАННЯ ЕКОНОМІЧНОЇ ЕФЕКТИВНОСТІ

Дипломна робота магістра присвячена дослідженню методів і засобів виявлення тональності текстової інформації для вдосконалення комп'ютерних систем. Важливим етапом виконання роботи є обґрунтування та розрахунок параметрів економічної ефективності від впровадження запропонованого методу і розробленого засобу. Тому у даному розділі проведено обчислення матеріальних витрат, затрат на заробітну плату інженера і керівника, а також інших параметрів, які впливають на економічну ефективність.

4.1. Розрахунок норм часу на виконання науково-дослідної роботи

Основні етапи при виконанні НДР можна визначити наступним чином:

- обґрунтування актуальності теми дипломної роботи магістра;
- аналіз наукових праць і практик реалізації методів і засобів визначення тональності текстової інформації;
- аналіз особливостей опрацювання відгуків користувачів при супроводі комп'ютерних систем;
- обґрунтування інтелектуальних методів аналізу та визначення тональності у відгуках користувачів про комп'ютерні системи;
- розробка інтелектуального методу виявлення тональності відгуків користувачів;
- розробка програмного засобу збору, аналізу та виявлення тональності текстової інформації при супроводі комп'ютерних систем.
- створення інструкції з використання розробленого засобу;
- оформлення інструкцій.

При оцінюванні тривалості виконання окремих робіт використовують нормативи часу або попередній досвід. До таких нормативів відносять тривалість написання операцій (команд), які для окремих підприємств становлять: для однієї операції – 0,5-1,6 год та 8 годин для п'яти операцій (тривалість зміни).

У разі їх відсутності звертаються до експертних оцінок по встановленню тривалості кожного етапу, яка при трьох оцінках обчислюється за формулою [31]

$$T_{\text{вс}} = (t_{\text{min}} + 4t_{\text{н.й}} + t_{\text{max}}) / 6, \quad (4.1)$$

При двох оцінках, експертна оцінка обчислюється за формулою:

$$T_{\text{вс}} = (3t_{\text{min}} + 2t_{\text{max}}) / 5, \quad (4.2)$$

де $T_{\text{вс}}$ – очікуване (середнє) значення тривалості виконання етапу (стадії);

t_{min} – мінімальна оцінка тривалості виконання етапу;

$t_{\text{н.й}}$ – найбільш імовірна оцінка тривалості виконання етапу;

t_{max} – максимальна оцінка тривалості виконання етапу.

Дані витрат часу на виконання окремих стадій (етапів) можна звести у табл. 4.1.

Основні етапи виконання НДР

№ та назва етапу	Середній час виконання стадії (етапу) інженером, год.	
	Інженер	Керівник
1. Обґрунтування актуальності теми дипломної роботи магістра	4	8
2. Аналіз наукових праць і практик реалізації методів і засобів визначення тональності текстової інформації	16	6
3. Аналіз особливостей опрацювання відгуків користувачів при супроводі комп'ютерних систем	12	2
4. Обґрунтування інтелектуальних методів аналізу та визначення тональності у відгуках користувачів про комп'ютерні системи	18	4
5. Розробка інтелектуального методу виявлення тональності відгуків користувачів	26	6
5 Розробка програмного засобу збору, аналізу та виявлення тональності текстової інформації при супроводі комп'ютерних систем	42	2
6 Створення інструкції з використання розробленого засобу	14	2
7 Оформлення інструкцій	4	2
Разом	136	32

Витрати часу керівника на виконання окремих стадій (етапів) при недостатній кількості інформації доцільно приймати в межах 5% сумарних витрат часу інженерів на виконання цих стадій (етапів).

4.2. Визначення витрат на оплату праці та відрахувань на соціальні заходи

Відповідно до Закону України «Про оплату праці» заробітна плата – це «винагорода, обчислена, як правило, у грошовому виразі, яку власник або

уповноважений ним орган виплачує працівникові за виконану ним роботу».

Розмір заробітної плати залежить від складності та умов виконуваної роботи, професійно-ділових якостей працівника, результатів його праці та господарської діяльності підприємства. Заробітна плата складається з основної та додаткової оплати праці.

Основна заробітна плата нараховується на виконану роботу за тарифними ставками, відрядними розцінками чи посадовими окладами і не залежить від результатів господарської діяльності підприємства.

Додаткова заробітна плата – це складова заробітної плати працівників, до якої включають витрати на оплату праці, не пов'язані з виплатами за фактично відпрацьований час. Нараховують додаткову заробітну плату залежно від досягнутих і запланованих показників, умов виробництва, кваліфікації виконавців. Джерелом додаткової оплати праці є фонд матеріального стимулювання, який створюється за рахунок прибутку.

Основна з/п складається із прямої з/п і доплати, яка при укрупнених розрахунках становить 25% – 35% від прямої з/п. При розрахунку з/п кількість робочих днів в місяці слід приймати – 21 дні/міс., що відповідає 168 год./міс. Розмір місячних окладів керівника та інженерів слід приймати згідно існуючих на даний час норм. Основна заробітна плата розраховується за формулою:

$$Z_{\text{осн.}} = T_c \cdot K_{\Gamma} \quad (4.3)$$

де T_c – тарифна ставка, грн.;

K_T – кількість відпрацьованих годин.

Посадові оклади (тарифні ставки) за розрядами Єдиної тарифної сітки визначаються шляхом множення окладу (ставки) працівника 1 тарифного розряду на відповідний тарифний коефіцієнт. У разі коли посадовий оклад (тарифна ставка) визначені у гривнях з копійками, цифри до 0,5 відкидаються, від 0,5 і вище – заокруглюються до однієї гривні. У 2019 році посадові оклади (тарифні ставки) розраховуються згідно з Законом України «Про Державний бюджет України на 2019 рік».

Мінімальна зарплата у 2019 р. складає 4173,00 грн., в погодинному розмірі 25,13 грн., прийmemo 100,00 грн. для інженера, для керівника – 150,00 грн.

Для інженера: $Z_{\text{осн.}} = 100,00 \cdot 136 = 13600,00$ грн.

Для керівника: $Z_{\text{осн.}} = 150,00 \cdot 32 = 4800,00$ грн.

Додаткова заробітна плата становить 10 – 15% від суми основної заробітної плати і обчислюється за формулою

$$Z_{\text{дод.}} = Z_{\text{осн.}} \cdot K_{\text{допл.}} \quad (4.4)$$

де $K_{\text{допл.}}$ – коефіцієнт додаткових виплат (0,1).

Для інженера: $Z_{\text{дод.}} = 1360,00 \cdot 0,1 = 1360,00$ грн.

Для керівника: $Z_{\text{дод.}} = 4800,00 \cdot 0,1 = 480,00$ грн.

Звідси загальні витрати на оплату праці ($V_{\text{оп.}}$) визначаються за формулою

$$V_{\text{оп.}} = Z_{\text{осн.}} + Z_{\text{дод.}} \quad (4.5)$$

Для інженера: $V_{\text{оп.}} = 13600,00 + 1360,00 = 14960,00$ грн.

Для керівника: $V_{\text{оп.}} = 4800,00 + 480,00 = 5280,00$ грн.

Таким чином загальна сума становить 20240,00 грн. Крім того, необхідно визначити відрахування на соціальні заходи:

- податок на доходи фізичних осіб: 18% – 3643,20 грн.;
- військовий збір: 1,5% – 303,60 грн.;
- єдиний внесок: 22% – 4452,80 грн.

У сумі зазначені відрахування становлять 41,5%. Отже, загальна сума відрахувань на соціальні заходи становитиме:

$$V_{C.3.} = \text{ФОП} \cdot 0,415 \quad (4.6)$$

де ФОП – фонд оплати праці, грн.

У даному випадку сума відрахувань становить: $V_{C.3.} = 20240,00 \cdot 0,415 = 8399,60$ грн.

Проведені розрахунки витрат на оплату праці зведемо у табл. 4.2.

Таблиця 4.2

Зведені витрати на заробітну плату

Категорія працівників	Основна заробітна плата, грн.			Додаткова заробітна плата, грн.	Нарахування на ФОП, грн.	Всього витрати на оплату праці, грн.
	Тарифна ставка, грн.	К-сть відпрацьованих	Фактично нарах. з/пл., грн.			
Інженер	100	136	13600,00	1360,00	6208,40	21168,40
Керівник проекту	150	32	4800,00	480,00	2191,20	7471,20
Разом			18400,00	1840,00	8399,60	28639,60

4.3. Розрахунок витрат на електроенергію

Затрати на електроенергію при використанні обладнання визначаються за формулою:

$$Z_e = W \cdot T \cdot S \quad (4.7)$$

де W – необхідна потужність, кВт;

T – кількість годин роботи обладнання;

S – вартість кіловат-години електроенергії.

Згідно з постановою НКРЕКП України від 05.10.2018 р. № 1177 вартість електроенергії становить 243,71 коп./кВт.год.

Потужність комп'ютера – 1 кВт, а кількість годин роботи обладнання згідно табл. 4.1 – 182 години.

Затрати на електроенергію становлять: $Z_e = 1 \cdot 168 \cdot 2,4371 = 409,43$ грн.

4.4. Розрахунок витрат на матеріали

Результати розрахунку затрат на матеріали наведено у табл. 4.3.

Таблиця 4.3.

Визначення величини затрат на матеріал

Найменування матеріальних ресурсів	Одиниця виміру	Норма витрат	Ціна за одиницю, грн.	Затрати матеріалів, грн	Транспортно – заготівельні витрати, грн.	Загальна сума витрат на матеріали, грн
Папір формату А4	пачка	2	100	200	-	200

Продовж. табл. 4.3

Найменування матеріальних ресурсів	Одиниця виміру	Норма витрат	Ціна за одиницю, грн.	Заграти матеріалів, грн	Транспортно – заготовельні витрати, грн.	Загальна сума витрат на матеріали, грн
Папір формату А1	шт.	10	6	60	-	60
Компакт диски	шт.	2	7	14	-	14
Разом						274

4.5. Розрахунок суми амортизаційних відрахувань

Характерною особливістю застосування основних фондів у процесі виробництва є їх відновлення. Для відновлення засобів праці у натуральному виразі необхідне їх відшкодування у вартісній формі, яке здійснюється шляхом амортизації.

Амортизація – це процес перенесення вартості основних фондів на вартість новоствореної продукції з метою їх повного відновлення.

Комп'ютери та оргтехніка належать до четвертої групи основних фондів. Для цієї групи річна норма амортизації дорівнює 60 % (квартальна – 15 %).

Для визначення амортизаційних відрахувань застосовуємо формулу:

$$A = \frac{B_B \cdot N_A}{100\%} \quad (4.8)$$

де A – амортизаційні відрахування за звітний період, грн.,

B_B – балансова вартість комп'ютера, на початок звітного періоду, грн.,

N_A – норма амортизації, яку приймемо на рівні 15%.

Амортизаційні відрахування при балансовій вартості ПК у 28000 грн. та нормі амортизації на рівні 15%, амортизаційні відрахування становитимуть:

$$A = \frac{28000 \cdot 15\%}{100\%} = 4200,00 \text{ грн.}$$

4.6. Обчислення накладних витрат

Накладні витрати пов'язані з обслуговуванням виробництва, утриманням апарату управління підприємства (фірми) та створення необхідних умов праці.

Накладні витрати можна встановити на рівні 20% від суми основної та додаткової заробітної плати працівників:

$$H_B = V_{OP} \cdot 0,2 \quad (4.9)$$

де H_B – накладні витрати, грн.,

V_{OP} – суми основної та додаткової заробітної плати працівників, грн..

У даному випадку накладні витрати становитимуть:

$$H_B = 20218,00 \cdot 0,2 = 4048,00 \text{ грн.}$$

4.7. Складання кошторису витрат та визначення собівартості науково-дослідних робіт

Собівартість (C_B) науково-дослідних робіт розрахуємо за формулою:

$$C_B = V_{o.p.} + V_{c.z} + Z_{m.v.} + Z_e + T_B + A + H_B \quad (4.10)$$

У даному випадку собівартість (СВ) науково-дослідних робіт розраховуємо за формулою: $C_B = 20240,00 + 8399,60 + 274,00 + 409,43 + 4200,00 + 4048,00 = 37571,03$ грн.грн.

Результати проведених вище розрахунків зведемо у табл. 4.4.

Таблиця 4.4.

Кошторис витрат на науково-дослідних робіт

Зміст витрат	Сума, грн.	В % до загальної суми
Витрати на оплату праці (основну і додаткову заробітну плату)	20240,00	53,87%
Відрахування на соціальні заходи	8399,60	22,36%
Матеріальні витрати	274	0,73%
Витрати на електроенергію	409,43	1,09%
Амортизаційні відрахування	4200,00	11,18%
Накладні витрати	4048,00	10,77%
Собівартість	37571,03	100,00%

4.8. Розрахунок ціни науково-дослідних робіт

Ціну науково-дослідних робіт можна визначити за формулою:

$$Ц = \frac{C_B \cdot (1 + P_{\text{рен.}}) + K \cdot B_{\text{н.і.}}}{K} \cdot (1 + \text{ПДВ}), \quad (4.11)$$

де $P_{\text{рен.}}$ – рівень рентабельності, 28 %;

K – кількість замовлень, од. (встановлюється лише при розробці програмного продукту та мікропроцесорних систем);

$B_{\text{н.і.}}$ – вартість носія інформації, грн. (встановлюється лише при розробці програмного продукту);

ПДВ – ставка податку на додану вартість, (20 %).

Оскільки розробка є прикладною, і використовуватиметься тільки для одного підприємства, то для розрахунку ціни не потрібно вказувати коефіцієнти K та $V_{н.і.}$, оскільки їх в даному випадку не потрібно.

Тоді, формула для обчислення ціни розробки буде мати вигляд:

$$Ц = C_B \cdot (1 + P_{рен.}) \cdot (1 + ПДВ) \quad (4.12)$$

Ціна становитиме: $Ц = 37571,03 \cdot (1 + 0,28) \cdot (1 + 0,2) = 57709,11$ грн.

Визначимо величину прибутку:

$$П = Ц - C_B \quad (4.13)$$

Прибуток буде становити: $П = 57709,11 - 37571,03 = 20138,07$ грн.

4.9. Визначення економічної ефективності і терміну окупності капітальних вкладень

Ефективність виробництва – це узагальнене і повне відображення кінцевих результатів використання робочої сили, засобів та предметів праці на підприємстві за певний проміжок часу.

Економічна ефективність (E_p) полягає у відношенні результату виробництва до затрачених ресурсів:

$$E_p = \frac{П}{C_B} \quad (4.14)$$

де $П$ – прибуток;

C_B – собівартість.

Економічна ефективність становить:

$$E_p = \frac{20138,07}{57709,11} = 0,54.$$

Поряд із економічною ефективністю розраховують термін окупності капітальних вкладень (T_p):

$$T_p = \frac{1}{E_p} \quad (4.15)$$

В даному випадку термін окупності становить: $T_p = \frac{1}{0,54} = 1,87$ року.

Про доцільність розробки програми можна сказати при врахуванні критеріїв, які наведені у табл. 4.5.

Таблиця 4.5.

Техніко-економічні показники НДР

№ з/п	Показник	Значення
1	Собівартість, грн	37571,03
2	Плановий прибуток, грн	20138,07
3	Ціна, грн	57709,11
4	Економічна ефективність	0,54
5	Термін окупності, рік	1,87

У результаті проведених розрахунків встановлено, що ціна розробленого методу і засобу визначення тональності текстової інформації при вдосконаленні комп'ютерних систем становить 57709,11 грн., а термін окупності 1,87 року. Тому доцільність такої розробки є економічно обґрунтованою.

РОЗДІЛ 5
ОХОРОНА ПРАЦІ ТА
БЕЗПЕКА В
НАДЗВИЧАЙНИХ
СИТУАЦІЯХ

5.1. Охорона праці

Метою дипломної роботи магістра є дослідження методів і засобів аналізу тональності текстової інформації при вдосконаленні комп'ютерних систем. Оскільки, усі етапи такого дослідження передбачають використання комп'ютерної та оргтехніки, то актуальним є аналіз та дотримання вимог з охорони праці і техніки безпеки при роботі з ПК.

Для забезпечення ефективності та оптимізації роботи аналітиків необхідно організувати безпечні умови праці. При цьому безпосередню відповідальність за порушення нормативно-правових актів з охорони праці несуть як самі аналітики тональності текстової інформації, так і їх керівники [23, 24].

Окрім цього, на робочих місцях осіб з оцінювання тональності тексту необхідно забезпечити дотримання вимог НПАОП 0.00-7.15-18 «Вимоги щодо безпеки та захисту здоров'я працівників під час роботи з екранними пристроями».

Приміщення, де розміщені робочі місця операторів, крім серверних, мають бути оснащені системою автоматичної пожежної сигналізації відповідно до вимог:

– переліку однотипних за призначенням об'єктів, які підлягають обладнанню автоматичними установками пожежогасіння та пожежної сигналізації, затвердженого наказом Міністерства України з питань надзвичайних ситуацій та у справах захисту населення від наслідків Чорнобильської катастрофи та представлених у НАПБ Б.06.004-2005;

– Державних будівельних норм "Інженерне обладнання будинків і споруд. Пожежна автоматика будинків і споруд", з димовими пожежними сповіщувачами та переносними вуглекислотними вогнегасниками.

В інших приміщеннях допускається встановлювати теплові пожежні сповіщувачі. Приміщення, де розміщені робочі місця аналітиків тональності текстової інформації, мають бути оснащені вогнегасниками, кількість яких визначається згідно з вимогами Типових норм належності вогнегасників, затверджених наказом Міністерства України з питань надзвичайних ситуацій та у справах захисту населення від наслідків Чорнобильської катастрофи від 02.04.2004 № 151 зареєстрованих у Міністерстві юстиції України 29.04.2004 за № 554/9153 (НАПБ Б.03.001-2004), і з урахуванням граничнодопустимих концентрацій вогнегасної рідини відповідно до вимог НАПБ А.01.001-2004.

Приміщення, в яких розміщуються робочі місця операторів сервера загального призначення, обладнуються системою автоматичної пожежної сигналізації та засобами пожежогасіння відповідно до вимог НАПБ Б.06.004-2005, ДБН В.2.5-13-98, НАПБ А.01.001-2004 і вимог нормативно-технічної та експлуатаційної документації виробника. Проходи до засобів пожежогасіння мають бути вільними.

Лінія електромережі для живлення комп'ютерів та периферійних пристроїв повинні бути виконаними як окрема групова трипровідна мережа шляхом прокладання фазового, нульового робочого та нульового захисного провідників. Нульовий захисний провідник використовується для заземлення (занулення) електроприймачів. Не допускається використовувати нульовий робочий провідник як нульовий захисний провідник. Нульовий захисний провідник прокладається від стійки групового розподільного щита, розподільного пункту до розеток електроживлення. Не допускається підключати на щиті до одного контактного затискача нульовий робочий та нульовий захисний провідники.

Площа перерізу нульового робочого та нульового захисного провідника в груповій трипровідній мережі має бути не менше площі перерізу фазового провідника. Усі провідники повинні відповідати номінальним параметрам мережі та навантаження, умовам навколишнього середовища, умовам розподілу провідників, температурному режиму та типам апаратури захисту, вимогам НПАОП 40.1-1.01-97.

У приміщенні, де одночасно експлуатуються понад п'ять комп'ютерів, на помітному та доступному місці встановлюється аварійний резервний вимикач, який може повністю вимкнути електричне живлення приміщення, крім освітлення.

Комп'ютери повинні підключатися до електромережі тільки за допомогою справних штепсельних з'єднань і електророзеток заводського виготовлення.

У штепсельних з'єднаннях та електророзетках, крім контактів фазового та нульового робочого провідників, мають бути спеціальні контакти для підключення нульового захисного провідника. Їхня конструкція має бути такою, щоб приєднання нульового захисного провідника відбувалося раніше, ніж приєднання фазового та нульового робочого провідників. Порядок роз'єднання при відключенні має бути зворотним.

Не допускається підключати комп'ютери до звичайної двопровідної електромережі, в тому числі – з використанням перехідних пристроїв.

Електромережі штепсельних з'єднань та електророзеток для живлення комп'ютерної техніки повинні бути виконаними за магістральною схемою, по 3-6 з'єднань або електророзеток в одному колі.

Штепсельні з'єднання та електророзетки для напруги 12 В та 42 В за своєю конструкцією мають відрізнятися від штепсельних з'єднань для напруги 127 В та 220 В. Штепсельні з'єднання та електророзетки, розраховані на напругу 12 В та 42 В, мають візуально (за кольором)

відрізнятися від кольору штепсельних з'єднань, розрахованих на напругу 127 В та 220 В.

При експлуатації програмної системи підтримки процесу оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем, важливим, з точки зору охорони праці, є забезпечення достатньої величини природного та штучного освітлення, які визначені у НПАОП 0.00-7.15-18.

Організація робочого місця аналітика з оцінювання тональності текстової інформації повинна забезпечувати відповідність усіх елементів робочого місця та їх розташування ергономічним вимогам "ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования".

Відстань від екрана до ока аналітиків, які працюють за комп'ютером визначається згідно з вимогами ДСанПіН 3.3.2.007-98.

Розміщення принтера на робочому місці має забезпечувати добру видимість екрана комп'ютера, зручність керування пристроєм введення-виведення інформації в зоні досяжності моторного поля згідно з вимогами ДСанПіН 3.3.2.007-98.

Таким чином, у результаті аналізу вимог щодо охорони праці користувачів комп'ютерів, визначено особливості організації робочих місць, вимог з електробезпеки, природного та штучного освітлення для ефективної і безпечної роботи аналітиків з оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем.

5.2. Запобігання забрудненню повітря, виробничих приміщень НХР, допустимі значення для основних речовин-забруднювачів; їх характеристика та засоби захисту

Небезпечні хімічні речовини використовуються як сировина (хлор для виготовлення хлорного вапна) чи допоміжний матеріал (бензол, який застосовується як розчинник) у виробництві різних промислових товарів. У деяких випадках вони є побічними продуктами, що створюються у

технологічних процесах. Тому важливим, з погляду захисту життя і здоров'я людини, є використання засобів запобігання забрудненню повітря та захисту виробничого персоналу.

Широкий спектр застосування небезпечних хімічних речовин у національній економіці сприяє значному розширенню виробництва та застосуванню їх в промисловості. При цьому значно збільшується їх асортимент: одержують багато нових хімічних сполук, які становлять небезпеку для оточуючого середовища і людей.

За фізіологічним впливом на організм людини всі шкідливі речовини поділяються на такі групи: подразнюючі, що вражають шляхи дихання, очі, шкіру, слизові оболонки (аміак, кислоти, сірчасті сполуки тощо); задушливі, які викликають токсичний набряк легень (сірководень, вуглекислий газ, метан, інертні гази, азот і т. д.); наркотичні, що спричиняють наркотичний вплив і впливають на центральну нервову систему (ацетон, бензин, леткі вуглеводні тощо); соматичні (миш'як, ртуть, свинець й ін.); канцерогенні речовини, що впливають, як правило, на злякисні новоутворення — пухлини (циклічні аміни, азбест, нікель, хром тощо).

Найбільш поширеними і небезпечними речовинами, що використовуються у промисловості і побуті, є аміак і хлор. Аміак використовується у промислових побутових холодильниках на м'ясокомбінатах, молокозаводах, овочевих базах, тобто там, де є необхідність в охолодженій продукції. При малих концентраціях він діє на людину збуджуючи, при великих – може призвести до інвалідності. Найкращі методи захисту в даних випадках – це застосування ізолюючого протигазу, респіратора, захисного костюма типу Л-1, гумових чобіт, рукавичок.

При виробництві або застосуванні хімічних речовин вони, потрапляючи у робочі приміщення чи безпосередньо на працівників, представляють небезпеку для здоров'я та нормальної життєдіяльності організму.

Умови зовнішнього середовища (наприклад, температура, вологість тощо) можуть посилювати чи послаблювати дію токсичних речовин. Так, висока вологість повітря посилює токсичні дії на організм соляної кислоти, фтористого водню.

Дія хімічних речовин може бути місцевою та загальною. Місцева дія зумовлюється опосередковано впливом дратівних речовин на тканини організму. Наприклад, мінеральні кислоти (соляна, азотна) та луки подразнюють шкіру. Загальна дія виявляється після потрапляння хімічних речовин у кров, причому деякі речовини діють на окремі органи, інші – викликають загальне отруєння організму.

За тривалістю дії шкідливі речовини ділять на три групи: летальні, що призводять до смерті (5% випадків) – термін дії до 10 діб; тимчасові, що призводять до нудоти, блювоти, набряку легенів, болю у грудях – термін дії від 2 до 5 діб; короткочасні – тривалість декілька годин (подразнення у носі, ротовій порожнині, головний біль, задуха, загальна слабкість).

Захист працівників від несприятливого впливу хімічних речовин здійснюється за допомогою таких заходів:

- удосконалення і розробки нових технологічних процесів, які виключають використання шкідливих хімічних речовин;
- заміни шкідливих речовин менш шкідливими (заміни метилового спирту бутиловим, жовтого фосфору – червоним при виробництві сірників);
- установлення концентрації хімічних речовин у сумішах (кількість миш'яку в кислотах для травлення металу не повинна перевищувати 0,2 %);
- комплексної механізації та автоматизації процесів, що супроводжуються шкідливими виділеннями;
- дистанційного управління технологічними процесами;
- раціонального планування цехів і обладнання (ізоляції шкідливих речовин);

- влаштування місцевої вентиляції для відсмоктування шкідливих речовин безпосередньо від місця їх утворення;
- використання індивідуальних засобів (спецодягу, окулярів, шоломів, масок, протигазів та респіраторів, антисептичних паст і т. д.);
- контролю за станом повітряного середовища на робочих місцях;
- токсикологічної експертизи і гігієнічної стандартизації всіх хімічних речовин.

5.3. Вплив іонізуючого випромінювання на користувачів комп'ютерних систем

Комп'ютерна та оргтехніка є джерелом практично усіх видів електромагнітного випромінювання (радіації). У залежності від впливу ці випромінювання бувають іонізуючими і неіонізуючими. До іонізуючого відноситься рентгенівське випромінювання, що широко використовується у медицині, до неіонізуючих - електромагнітне поле (випромінювання) наднизької і низької частоти.

Іонізуюче випромінювання, впливаючи на об'єкт, зокрема, на клітини людини, викликає їхнє ушкодження за рахунок утворення іонів. Ці ушкодження можуть бути летальними, коли клітина гине, і сублетальними, коли клітина виживає, але інформація, "защита" у ній, псується. Такі клітини можуть бути джерелом виникнення раку.

Електромагнітні випромінювання наднизької частоти не мають здатності викликати іонізацію, а, відповідно, і мутації. Їхня дія на живу клітину мало вивчена, однак відомо, що вони за відносно короткий термін впливу (10-15 років) не призводять до виникнення злоякісних пухлин. Існує величезна кількість досліджень електромагнітного поля наднизької частоти, одні з яких доводять, що цей вид випромінювання шкідливий для здоров'я, а інші - зворотне. Усі роботи, що доводять шкоду електромагнітного поля наднизької частоти, спираються на епідеміологічні

дані. Це означає, що тут можуть бути неточності, не виключені впливи інших факторів. Конкретного, що ушкоджує, механізму впливу електромагнітного поля наднизької частоти ніхто не знає.

Ефекти, викликані дією іонізуючих випромінювань (радіації), систематизуються за видами ушкоджень і часом прояву. Ефекти за видами ушкоджень класифікуються на 3 групи: соматичні, соматико-стахотичні (випадкові, ймовірні), генетичні. Час прояву вказує дві групи поразок – ранні (або гострі) і пізні. Ранні поразки бувають тільки соматичні. Це призводить до смерті або променевої хвороби. Постачальником таких часток є в основному ізопопи, що мають коротку тривалість життя, - випромінювання, потік нейтронів.

Дії електромагнітних випромінювань можна усі поділити в такий спосіб: іонізуючі випромінювання, такі, як рентгенівське, при визначеній дозі опромінення можуть викликати виникнення злоякісних пухлин. Електромагнітні поля наднизької частоти не складають загрози для здоров'я людини, однак, у силу того, що їхня дія мало вивчена, рекомендується зменшити чи звести до мінімуму зустріч з ними .

Рентгенівське випромінювання, що виходить від монітора, мізерно мале і порівняно з природним радіаційним фоном. Це означає те, що людина, яка сидить поруч з дисплеєм чи гуляє по вулиці - одержує дозу приблизно ту саму. Виключення складають браковані монітори, вберегтися від яких можна, вибираючи відому марку, у відомого постачальника й у відомому магазині.

За способом захисту людини від ураження електричним струмом відеотермінали, ЕОМ, периферійні пристрої ЕОМ та устаткування для обслуговування, ремонту та налагодження ЕОМ повинні відповідати I класу захисту згідно з ГОСТ 12.2.007.0 "ССБТ. Изделия электротехнические. Общие требования безопасности" та ГОСТ 25861-83 "Машины вычислительные и системы обработки данных. Требования электрической и

механической безопасности и методы испытаний" або повинні бути заземлені відповідно до ДНАОП 0.00-1.21-98.

Є неприпустимим використання клем функціонального заземлення для підключення захисного заземлення.

Вимоги щодо допустимих значень неіонізуючого електромагнітного випромінювання: напруженість електромагнітного поля на відстані 50 см навкруги ВДТ за електричною складовою не повинна перевищувати:

- у діапазоні частот 5 кГц - 2 кГц - 25 В/м,
- у діапазоні частот 2 кГц - 400 кГц - 2,5 В/м;
- щільність магнітного потоку не повинна перевищувати:
- у діапазоні частот 5 кГц - 2 кГц - 250 нТл,
- у діапазоні частот 2 кГц - 400 кГц - 25 нТл;
- поверхневий електростатичний потенціал не повинен перевищувати 500 В;
- потужність дози рентгенівського випромінювання на відстані 5 см від екрану та інших поверхонь ВДТ не повинна перевищувати 100 мкР/год.

Дотримання та забезпечення заходів щодо контролю та захисту від іонізуючого випромінювання, спричиненого комп'ютерною та оргтехнікою, дає змогу знизити негативний вплив на здоров'я людей і підвищити якість життя. Тому дотримання вимог нормативних документів є обов'язковими для виконання при використанні сучасних електронно-обчислювальних засобів.

РОЗДІЛ 6 ЕКОЛОГІЯ

6.1. Зниження енергоємності та енергозбереження

Ресурсо- і енергозбереження та оздоровлення навколишнього природного середовища – це дві сторони єдиного процесу досягнення еколого-економічної збалансованості у функціонуванні національної економіки. Заходи економії паливно-енергетичних ресурсів (ПЕР) характеризуються тією перевагою, що дозволяють зменшити викиди всіх шкідливих речовин. Слід зазначити, що за допомогою інших природоохоронних заходів викиди однієї речовини можуть замінитися викидами іншої речовини. Захист навколишнього середовища за допомогою енергозбереження не пов'язаний із додатковими витратами енергії, тоді як інші природоохоронні методи часто передбачають суттєве збільшення енергоспоживання [28].

Екологічний (природоохоронний) ефект енергозбереження обмежується не лише сферою використання енергії. Кожна заощаджена при використанні тонна палива, гікалорія теплоти чи заощаджена кіловат-година електроенергії дозволяють також забезпечити значний природоохоронний ефект на всіх попередніх енергетичних стадіях, пов'язаних із видобуванням палива, його збагаченням, переробкою та транспортуванням; виробництвом, транспортуванням електричної та теплової енергії до споживача та її розподілом.

Упровадження заходів з енергозбереження надає більш широкі можливості для виконання міжнародних обов'язків України, зокрема «Концепції про транскордонне забруднення повітря на великі відстані», яку наша держава ратифікувала у 1980 р.

Учені винаходять новітні конструкції і прилади. Новітні технології роблять наше життя більш комфортним, спрощують нам роботу і певні дії.

Вони роблять відкриття в області медицини і багатьох інших галузях. Так само їм вдалося зробити відкриття в галузі альтернативних джерел енергії, які дають змогу скоротити витрати ресурсів та енергії на виробництво товарів, продукції чи надання послуг [28].

До складу комп'ютера входить ряд пристроїв, які при роботі споживають велику кількість енергії. Такими компонентами є монітори, жорсткі диски, системна плата, процесор і деякі інші пристрої.

Для керування режимами електроживлення та мінімізації пов'язаних з цим процесом втрат електроенергії розроблено кілька стандартних технологій.

Сучасні комп'ютери розробляються відповідно до положень сертифіката технології Energy Star агентства по захисту навколишнього середовища EPA (Environmental Protection Agency).

Для отримання зазначеного сертифікату параметри операційної системи, I/O та BIOS повинні відповідати вимогам стандарту енергозбереження. Так, наприклад, системна плата забезпечується мікросхемою BIOS (цей компонент входить до складу чіпсета), що підтримує технологію ACPI Plug and Play, а також функції SMBIOS (System Management BIOS) [28].

Для здійснення контролю і управління енергозберігаючими функціями комп'ютера компоненти системної плати включають елементи реалізації протоколу управління апаратним забезпеченням WfM (Wired for Management). Специфікація цього стандарту передбачає підтримку системною платою наступних функцій:

- виявлення відкритого системного блоку;
- моніторингу роботи вентиляторів і управління швидкісними режимами роботи вентиляторів;
- моніторинг температурних режимів і постійних напруг живлення елементів системного блоку.

Система моніторингу параметрів апаратних засобів, швидкісних

режимів роботи вентиляторів та управління вентиляторами базується на мікросхемі ASIC, а також наступних апаратних вузлах: датчики напруги живлення, контролюють постійні напруги +5 В, +12 В, +3,3 В, +5 У SB , щоб вони не виходили за межі поля допуску; віддалений датчик температури поблизу апаратних засобів на системній платі; датчики, що контролюють температуру процесора і температурний режим навколо нього; датчик навколишнього середовища компонент ASIC апаратного моніторингу та управління вентиляторами [28].

Система управління, побудована на базі ASIC дозволяє регулювати швидкість обертання вентиляторів , а в разі необхідності відключати і включати їх, що забезпечує реалізацію процесу енергозбереження при переході ПК в режим очікування [28].

Таким чином, зниження енергоємності ПК і як наслідок забезпечення енергозбереження базується на системах управління живлення компонентів комп'ютера. Основне призначення будь-якої системи управління живленням – автоматично переводити комп'ютер або окремі його пристрої в один з режимів (станів) зниженого енергоспоживання, що дає змогу забезпечити низьке споживання напруги живлення.

6.2. Вимоги до мікроклімату, вмісту аеронів і шкідливих хімічних речовин у повітрі приміщень експлуатації моніторів і ПЕОМ. Шляхи дотримання цих вимог

Висока чи низька температура повітря в приміщенні з ПК негативно впливає на функціональний стан користувача. Недостатня вологість в приміщенні призводить до надмірного висихання слизових оболонок очей, носа, горла та до нагромадження зарядів статичної електрики, що утворюються в процесі роботи ПК [29].

Разом з тим недопустима вологість повітря більше 75%. На робочих місцях користувачів ПК параметри мікроклімату мають відповідати

ВИМОГАМ

ДСН 3.3.6.042-99 “Санітарні норми мікроклімату виробничих приміщень”,
ДСан Пін 3.3.2-007-98, ДНАОП 0.00-1.31-99.

У табл. 6.1 приведені норми мікроклімату для приміщень з ВДТ ЕОМ та ПЕОМ (ПК).

Таблиця 6.1

Норми мікроклімату для приміщень з ВДТ ЕОМ та ПЕОМ

Пора року	Категорія робіт**	Температура повітря, °С	Відносна вологість повітря, %	Швидкість руху повітря, м/с
Холодна	легка-1а	22-24	60-40	0,1
легка-1б	21-23	60-40	0,1	
Тепла	легка-1а	23-25	60-40	0,1
легка-1б	22-24	60-40	0,2	

Холодна пора року – період року, який характеризується середньодобовою температурою зовнішнього повітря, що дорівнює $+10^{\circ}\text{C}$ і нижче.

Тепла пора року – період року, який характеризується середньодобовою температурою зовнішнього середовища вище $+10^{\circ}\text{C}$.

До категорії 1а належать роботи, що виконуються сидячи і не потребують фізичного напруження, при яких витрата енергії дорівнює 105-140Вт (90-120 ккал/год.).

До категорії 1б належать роботи, що виконуються сидячи, стоячи або пов'язані з ходінням та супроводжуються деяким фізичним напруженням, при яких витрата енергії дорівнює 141-175 Вт (121-150 ккал/год.).

Повітря яким дихають користувачів ПК складається з позитивно та негативно заряджених часток-іонів. Давно відомо про цілющий вплив на

людину негативних іонів. Вони підвищують рівень гаммаглобуліна в крові, що збільшує опір організму хвороботворним бактеріям, а також стимулює вироблення бетаендорфіна, що інколи називають “гормоном щастя”, який позитивно впливає на настрій, запобігаючи депресіям та підвищуючи працездатність [29].

При роботі з ВДТ і ПК іонний склад повітря на робочому місці користувача змінюється. Уже через 5 хвилин роботи кількість легких іонів знижується у 8 разів, а через 3 години є майже нульовою. Також зменшується концентрація середніх та важких негативно заряджених часток, а концентрація позитивних іонів зростає, що шкідливо впливає на здоров'я користувача комп'ютера [29].

Такий вплив проявляється в зниженні його працездатності (швидка втомлюваність, головний біль), в погіршенні короткочасової пам'яті, діяльності серцево-судинної системи, бронхо-легеневого апарату (збільшення частоти пульсу та дихання (із-за недостатнього поступлення кисню в кров)).

У приміщеннях з комп'ютерами має бути забезпечений 3-кратний обмін повітря за годину. Адже під час розумової праці мозок людини споживає в 9-10 разів більше кисню, а ПК його забирає, виникає кисневе голодування.

Для забезпечення нормованого мікроклімату та рівня іонізації повітря на робочих місцях користувачів рекомендується застосовувати припливно-втяжну вентиляцію чи систему кондиціонування повітря, прилади зволоження та/або установки генерації негативних іонів (аероіонізатори). Знизити деіонізацію повітря в зоні дихання користувача ПК дозволяє також встановлений на монітор захисний заземлений екран [29].

На сьогоднішній день деякі компанії починають випускати монітори із влаштованими іонізаторами повітря. Так, наприклад, компанія Samsung Electronics заявила, що почала продавати РК-монітори із вмонтованими іонізаторами повітря. При включенні іонізатора він починає випромінювати

негативно заряджені іони, які перешкоджають електризації пилу та скупченню бактерій біля робочого місця користувача. Сучасні модифіковані аероіонізатори дозволяють одночасно очищувати і іонізувати повітря негативними іонами.

У великих приміщеннях можна використовувати аероіонізатори типу люстри Чижевського.

Вимоги до вмісту шкідливих речовин у повітрі робочої зони, відповідно до ГОСТ 12.1.005-88, наступні:

- концентрація озону – не більше 0,1 мг/м³;
- вміст оксидів азоту – не більше 5 мг/м³;
- вміст пилу – не більше 4 мг/м³.

Таким чином, дотримання вимог щодо вмісту аеронів, застосування засобів іонізації повітря та знешкодження шкідливих речовин у повітрі дають змогу дотримуватись рекомендацій стандартів щодо мікроклімату у виробничих приміщеннях з використанням ПЕОМ і тим самим забезпечити оптимальні умови праці користувачів ПК.

ВИСНОВКИ

Основні наукові та практичні результати роботи полягають у наступному:

1. Проведено аналіз відкритих та спеціалізованих джерел одержання інформації для вдосконалення комп'ютерних систем у результаті якого визначено шляхи збору та аналізу текстових даних, їх інтерпретації та подальшого прийняття рішень щодо модернізації складових компонентів або системи в цілому.

2. Проаналізовано сучасний стан розвитку методів і засобів комп'ютерної лінгвістики на різних рівнях представлення текстової інформації та обгрунтовано необхідність застосування комплексного підходу щодо виявлення та аналізу тональності текстової інформації на синтаксичному, лексичному та семантичному рівнях.

3. Досліджено прикладні аспекти застосування комп'ютерної лінгвістики, що дало змогу визначити та обгрунтувати методи добування та аналізу відгуків користувачів про властивості комп'ютерних систем і її поведінки на основі аспектного підходу.

4. Формалізовано представлення відгуків користувачів і задач, які при цьому необхідно розв'язати, у вигляді квантилю (сутність-аспект-тональність-користувач-час), що дало змогу врахувати позитивні і негативні властивості, реалізовані у комп'ютерних системах та в подальшому відобразити їх при вдосконаленні систем.

5. Обгрунтовано необхідність застосування частотних методів при виявленні аспектів комп'ютерних систем, що позначаються на враженні кінцевих користувачів і дають змогу виявити найбільш важливі властивості комп'ютерних систем.

6. Проведено аналіз методів добування аспектів на основі підходу побудови правил, машинного навчання і тематичного моделювання, у результаті якого обгрунтовано необхідність використання методів

машинного навчання в комплексі з частотними методами синтаксичного аналізу при вирішенні задач вдосконалення комп'ютерних систем.

7. Обґрунтовано застосування лінгвістично-інженерного підходу при оцінюванні тональності текстової інформації щодо вдосконалення комп'ютерних систем, що дає змогу врахувати якість аспектів на основі слів, що їх оцінюють та відстані між аспектами і такими словами.

8. Проаналізовано методи і метрики оцінювання тональності текстової інформації і встановлено доцільність застосування F-міри, що враховує повноту і точність оцінювання тональності відгуків користувачів і дає змогу виявити найбільш проблемні властивості комп'ютерної системи.

9. При розробці системи виявлення та аналізу тональності текстової інформації для вдосконалення комп'ютерних систем побудовано діаграми варіантів використання, що дало змогу визначити функціональні вимоги до системи та ролі користувачів.

10. Розроблено схему бази даних для зберігання та маніпулювання даними щодо оцінювання тональності текстової інформації (відгуків користувачів), що дало змогу забезпечити гнучкість та масштабованість процесу вдосконалення комп'ютерних систем.

11. Спроектовано архітектуру системи виявлення та оцінювання тональності відгуків користувачів про комп'ютерну систему або її компоненти, що дало змогу реалізувати програмні модулі керування параметрами алгоритмів і моделей машинного навчання.

12. Обґрунтовано економічну доцільність щодо впровадження запропонованих методів і засобу виявлення та оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем шляхом обчислення показників економічної ефективності, зокрема собівартості, ціни і терміну окупності, який становить 1,87 року.

13. Проаналізовано вимоги з охорони праці та безпеки в надзвичайних ситуаціях, що дало змогу визначити шляхи мінімізації негативного впливу комп'ютерної техніки на користувачів програмного засобу виявлення та

оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем.

14. Проаналізовано методи і заходи щодо зниження енергоємності та енергозбереження, а також проведено аналіз вимог до мікроклімату у приміщеннях, де експлуатуються ПК.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ермаков А.Е., Киселев С.Л. Лингвистическая модель для компьютерного анализа тональности публикаций СМИ //Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог. 2005. С. 282-285.
2. Борисова Е. Г., Пирогова Ю. К. Моделирование нетривиальных условий понимания сообщения (на примере иронии) //Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» 2013. С. 148-162.
3. Алексеева С. В., Кольцов С. Н., Кольцова О. Ю. Linis-crowd. org: лексический ресурс для анализа тональности социально-политических текстов на русском языке //Компьютерная лингвистика и вычислительные онтологии: сборник научных статей. Труды XVIII объединенной конференции «Интернет и современное общество» (IMS-2015), Санкт-Петербург, 2015. С. 25-34.
4. Воронцов К.В., Потапенко А.А. Модификации EM-алгоритма для вероятностного тематического моделирования // Машинное обучение и анализ данных, 2013. V.1. № 6. С. 657–686.
5. Лукашевич Н. В., Рубцова Ю. В. Объектно-ориентированный анализ твитов по тональности: результаты и проблемы //Аналитика и управление данными в областях с интенсивным использованием данных. 2015. С. 499-507.
6. Пазельская А. Г., Соловьев А. Н. Метод определения эмоций в текстах на русском языке //Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог»(Бекасово, 25–29 мая 2011 г.). М.: Изд-во РГГУ. 2011. С. 17.
7. Марчук А. А., Уланов А. В., Макеев И. В., Чугреев, А. А. Автоматическое извлечение параметров продуктов из текстов отзывов при помощи интернет-статистик // Труды Международной конференции

Автоматическая обработка текстов и анализ данных Компьютерная лингвистика и информационные технологии Диалог- 2013. 2013. т.2. С. 81–91.

8. Рубцова Ю. В. Метод построения и анализа корпуса коротких текстов для задачи классификации отзывов //Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды XV Всероссийской научной конференции RCDL. 2013. С. 269-275.

9. Русначенко Н., Лукашевич Н. Методы интеграции лексиконов в машинной обучение для систем анализа тональности // Искусственный интеллект и принятие решений, N 2. 2017. С. 78-89.

10. Рубцова Ю. В. Построение корпуса текстов для настройки тонового классификатора //Программные продукты и системы. 2015. №. 1 (109).

11. Телия В.Н. Коннотативный аспект семантики номинативных единиц. — М.: Наука, 1986. — 143 с.

12. Andrzejewski D., Zhu X., Craven M. Incorporating domain knowledge into topic modeling via Dirichlet forest priors // Proceedings of ICML. 2009. P. 25–32.

13. Amigo E., Albornoz J.C., Chugur I., Corujo A., Gonzalo J., Martin T., Meij E., de Rijke M, Spina D. Overview of RepLab 2013: Evaluating online reputation monitoring systems //Information Access Evaluation. Multilinguality, Multimodality, and Visualization. Springer Berlin Heidelberg, 2013. P. 333

14. Arkhipenko K., Kozlov I., Trofmovich J., Skorniakov K., Gomzin A., Turdakov D. Comparison of Neural Network Architectures for Sentiment Analysis of Russian Tweets // In Proceedings of International Conference on computational linguistics and intellectual technologies Dialog-2016. 2016. P. 50-58.

15. Amigo E., Corujo A., Gonzalo J., Meij E., Rijke M. Overview of RepLab 2012: Evaluating Online Reputation Management Systems // CLEF-2012.

16. Aue A., Gamon M. Customizing sentiment classifiers to new domains: A case study // In Proceedings of International Conference on Recent Advances in Natural Language Processing, Borovets, BG, 2005.
17. Baccianella, S., Esuli, A., Sebastiani F. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining // Proceedings of LREC-2010, V. 10, 2010. P. 2200-2204.
18. Bagheri A., Saraee M., de Jong F. An Unsupervised Aspect Detection Model for Sentiment Analysis of Reviews // Natural Language Processing and Information Systems. Springer: Berlin Heidelberg, 2013. P. 140–151.
19. Ben-Ami Z., Feldman R., Rosenfeld B. Entities' Sentiment Relevance //In Proceedings of ACL-2014. 2014. P. 87-92.
20. Benamara F., Taboada M., Mathieu Y. Evaluative language beyond bags of words: Linguistic insights and computational applications //Computational Linguistics, V.43, 2017. P. 201-264.
21. Blinov P.D., Kotelnikov E.V. Semantic Similarity for Aspect-Based Sentiment Analysis // Proceedings of International Conference of Computational Linguistics and Intellectual Technologies Dialog-2015. 2015. V. 2. P. 23–33.
22. Bollen J., Mao H., Zeng X. Twitter mood predicts the stock market //Journal of computational science. 2011. T. 2. N. 1. P. 1-8.
23. Blei D., Ng A., Jordan M. Latent dirichlet allocation // The Journal of Machine Learning Research, 2003. № 3. P. 993–1022.
24. Blair-Goldensohn S., Hannan K., McDonald R., Neylon T., Reis G. A., Reynar J. Building a sentiment summarizer for local service reviews //Proceedings of WWW Workshop on NLP in the Information Explosion Era. 2008.
25. Cambria E., Livingstone A., and A. Hussain. The hourglass of emotions // Cognitive Behavioural Systems, Lecture Notes in Computer Science, vol. 7403, Springer, 2012, P. 144–157.

26. Bradley M.M., Lang P.J. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical report C-1, the center for research in psychophysiology, University of Florida, 1999. P. 1-45.

27. Фаулер М. Архитектура корпоративных программных приложений. М.: Издательский дом "Вильямс". 2006. 544 с.

28. Жидецкий В. Охорона праці користувачів комп'ютерів. Львів: Афіша, 2000. 176 с.

29. Желібо Є. Безпека життєдіяльності. К.: 2001. 483 с.

30. Джигирей В. Екологія та охорона навколишнього природного середовища. Навчальний посібник. К.: Знання, 2000. 256 с.

Додаток А
Тексти наукових публікації

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Тернопільський національний технічний університет імені Івана Пулюя (Україна)
Національна академія наук України
Університет імені П'єра і Марії Кюрі (Франція)
Маріборський університет (Словенія)
Технічний університет у Кошице (Словаччина)
Вільнюський технічний університет ім. Гедимінаса (Литва)
Шяуляйська державна колегія (Литва)
Жешувський політехнічний університет ім. Лукасевича (Польща)
Білоруський національний технічний університет (Республіка Білорусь)
Міжнародний університет цивільної авіації (Марокко)
Національний університет біоресурсів і природокористування України (Україна)
Наукове товариство ім. Шевченка
ГО «Асоціація випускників Тернопільського національного технічного університету імені Івана Пулюя»

**АКТУАЛЬНІ ЗАДАЧІ
СУЧАСНИХ ТЕХНОЛОГІЙ**

Збірник

тез доповідей

Том II

**VIII Міжнародної науково-технічної
конференції молодих учених та студентів**

27-28 листопада 2019 року



**УКРАЇНА
ТЕРНОПІЛЬ – 2019**

ЗМІСТ

СЕКЦІЯ: КОМП'ЮТЕРНО-ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА СИСТЕМИ ЗВ'ЯЗКУ

1.	М.М. Баранчук, А.М.Шельвіка, П.Д. Стухляк РОЗРОБКА ТА ОПТИМІЗАЦІЯ СИСТЕМИ УПРАВЛІННЯ ТЕХНОЛОГІЧНИМ ПРОЦЕСОМ ВИРОБНИЦТВА КВАСУ	5
2.	Д.О. Батошний, А.П. Петрук, Р.З. Золотий РОЗРОБКА ТА ОПТИМІЗАЦІЯ СИСТЕМИ АНАЛІЗУ КЛІМАТИЧНИХ ДАНИХ	6
3.	М.С. Бедрійчук АДАПТИВНИЙ МЕТОД ВИБОРУ КАНАЛУ ЗВ'ЯЗКУ ДЛЯ РОЗУМНОГО БУДИНКУ	7
4.	С.В. Бенедюк, Б.І. Яворський МЕТОД ВИЯВЛЕННЯ КОРИСНОГО СИГНАЛУ У ШУМІ В КОРОТКОХВИЛЬОВОМУ ДІАПАЗОНІ РАДІОХВИЛЬ	8
5.	Є. М. Білоус, С. П. Галайко, А. А. Липак, А. О. Порядок, Н. В. Цвіркун ДОСЛІДЖЕННЯ ВПЛИВУ ВІДХИЛЕНЬ НЕСУЧОЇ ПЛАТФОРМИ НА ЗМІЩЕННЯ ДІАГРАМИ НАПРАВЛЕНОСТІ АНТЕНИ	9
6.	А.Р. Бориславський ВИЯВЛЕННЯ ШКІДЛИВОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ В ОС ANDROID НА ОСНОВІ ІНТЕЛЕКТУАЛЬНИХ ТЕХНОЛОГІЙ	10
7.	Р.А. Буцій ВПРОВАДЖЕННЯ БЕЗДРОТОВИХ LoRa MESH-МЕРЕЖ НА БАЗІ ПЛАТФОРМИ ARDUINO UNO В СИСТЕМИ ІоТ	11
8.	В.В. Вайман ВИКОРИСТАННЯ QR-КОДУ В СУЧАСНОМУ СВІТІ	12
9.	Д.А. Войцехівський, С.І. Глазков, В.О. Наумов, І.Г.Добротвор ДОСЛІДЖЕННЯ АВТОМАТИЗОВАНИХ ПРОЦЕСІВ УТИЛІЗАЦІЇ ТА ПЕРЕРОБКИ ПЛАСТИКОВИХ ВИРОБІВ	13
10.	С.Б. Волох, Р.М. Кирилів, Д.І. Полоз, І.В. Півторак, Ю.О. Апостол ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ ДОСЛІДЖЕННЯ ВПЛИВУ НЕКОРЕЛЬОВАНОЇ ЕЛІПТИЧНОСТІ РОТОРА І СТАТОРА НА ВИНИКНЕННЯ ВІБРАЦІЙ ЕЛЕКТРОДВИГУНА	14
11.	Р.Р. Гавап, В.В. Яцишин ПІДХОДИ ДО ВДОСКОНАЛЕННЯ КОМП'ЮТЕРНИХ СИСТЕМ НА ОСНОВІ АНАЛІЗУ ТОНАЛЬНОСТІ ВІДГУКІВ КОРИСТУВАЧІВ	16

УДК 004.89

Р.Р. Гаван, В.В. Яцишин, канд. техн. наук, доц.

Тернопільський національний технічний університет імені Івана Пулюя, Україна

ПІДХОДИ ДО ВДОСКОНАЛЕННЯ КОМП'ЮТЕРНИХ СИСТЕМ НА ОСНОВІ АНАЛІЗУ ТОНАЛЬНОСТІ ВІДГУКІВ КОРИСТУВАЧІВ

R.R. Gavan, V.V. Yatsyshyn PhD, Assoc. Prof.

APPROACHES OF COMPUTER SYSTEMS IMPROVEMENT BASED ON SENTIMENT ANALYSIS OF USER'S FEEDBACKS

Сучасні інформаційні технології характеризуються розвитком і застосуванням нових підходів, моделей та методів проектування комп'ютерних систем, що дає змогу в короткі терміни із заданою якістю та ефективністю реалізувати готові продукти та програмно-апаратні комплекси. Особливістю використовуваних моделей і методів проектування комп'ютерних систем є залучення зацікавлених осіб та користувачів до процесу їх розробки, супроводу та підтримки. Враховуючи розвинутість соціальних мереж загального використання, віртуальних професійних об'єднань, наявність тематичних блогів та систем електронної комерції з можливістю формування відгуків про придбаний товар чи послугу, зростає кількість інформації, яка є важливою для розробників і дозволяє визначити шляхи вдосконалення їхньої продукції. Для вдосконалення та підвищення якості комп'ютерних систем необхідно впроваджувати методи і засоби штучного інтелекту для автоматизованого визначення тональності відгуків користувачів. На основі аналізу відгуків користувачів можна визначити недоліки та переваги в існуючих комп'ютерних системах, виявити можливі шляхи їх вдосконалення і побудувати моделі модернізації при випуску нових версій продукту.

Оскільки, відгуки користувачів про комп'ютерні системи представляються, зазвичай, у тестовому вигляді із можливим застосуванням різноманітних емограм, тому актуальними задачами при вдосконаленні комп'ютерних систем є розробка, обґрунтування і застосування методів і засобів аналізу текстової інформації та емограм. Проведемо аналіз існуючих підходів до виявлення тональності у текстовій інформації. Виявлення тональності текстової інформації, або по іншому класифікації тексту за класами «позитивний», «негативний», «нейтральний» сьогодні досліджується на трьох рівнях: рівень документа (Document level); рівень речення (Sentence level); рівень сутностей та властивостей (Entity and Aspect level).

Основна задача, яка розв'язується при аналізі семантики на рівні документа, полягає у класифікації цілого документа і визначенні його приналежності до класу «позитивний» чи «негативний». Особливістю даного рівня є те, що кожен документ описує лише одну сутність і не передбачає порівняння з іншими. Основною одиницею дослідження на рівні речення є власне саме речення. При цьому речення може належати лише до одного з класів – «позитивний», «негативний», «нейтральний». Класифікацію на рівні речення часто пов'язують із суб'єктивною класифікацією, що передбачає визначення різниці між об'єктивними реченнями та суб'єктивними, у яких виражені думки про певний об'єкт або його властивість. На відміну від аналізу на рівні документа і на рівні речення, рівень сутностей та атрибутів (аспектів) дає змогу оперувати настроями і думками користувачів про товар чи систему. Даний рівень основну увагу концентрує не на синтаксичних конструкціях, а на проявах позитиву чи негативу у висловлюваннях. Тому його застосування при вдосконаленні комп'ютерних систем є найбільш обґрунтованим.

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ІВАНА ПУЛЮЯ**

МАТЕРІАЛИ

VII НАУКОВО-ТЕХНІЧНОЇ КОНФЕРЕНЦІЇ

**«ІНФОРМАЦІЙНІ МОДЕЛІ,
СИСТЕМИ ТА ТЕХНОЛОГІЇ»**



11–12 грудня 2019 року

**ТЕРНОПІЛЬ
2019**

СЕКЦІЯ 3. КОМП'ЮТЕРНІ СИСТЕМИ ТА МЕРЕЖІ	
В. Владика, Д. Величко, Г. Осухівська ТЕХНОЛОГІЇ ПЕРЕДАВАННЯ ДАНИХ В СИСТЕМІ «ЦИФРОВА ЛІКАРНЯ»	109
В. Барбарич, Ю. Івануса ШКІДЛИВИЙ ВПЛИВ WI-FI НА ОРГАНІЗМ ЛЮДИНИ	110
М. Бедрійчук РЕАЛІЗАЦІЯ МЕТОДУ АДАПТИВНОЇ СЕЛЕКЦІЇ КАНАЛІВ ЗВ'ЯЗКУ ДЛЯ СИСТЕМИ «РОЗУМНИЙ БУДИНОК»	112
Д. Войтина, В. Яцишин АНАЛІЗ ОСОБЛИВОСТЕЙ КОМП'ЮТЕРНИХ СИСТЕМ КРИТИЧНОГО ПРИЗНАЧЕННЯ	113
Р. Гаван, В. Яцишин ОСОБЛИВОСТІ АНАЛІЗУ ТОНАЛЬНОСТІ ТЕКСТУ НА ОСНОВІ АСПЕКТІВ ПРИ ВДОСКОНАЛЕННІ КОМП'ЮТЕРНИХ СИСТЕМ	114
С. Галап, В. Яцишин ОСОБЛИВОСТІ ПРОЕКТУВАННЯ «РОЗУМНИХ СИСТЕМ» З МОЖЛИВІСТЮ ВЗАЄМОДІЇ З ХМАРНИМИ СЕРВІСАМИ	115
Р. Гайдук, Д. Михалик РОЗРОБКА СОЦІАЛЬНОЇ МЕРЕЖІ ДЛЯ БІБЛІОТЕК З ВИКОРИСТАННЯМ .NET ТЕХНОЛОГІЙ	116
Ю. Голояд МЕТОДИ ТА АЛГОРИТМИ РОЗПІЗНАВАННЯ ОБРАЗІВ	117
І. Голуб, О. Ясній МЕТОДИ ТА ЗАСОБИ ПОБУДОВИ МЕРЕЖЕВИХ КОМУТАТОРІВ ЗПІДТРИМКОЮ ТЕХНОЛОГІЙ GERON ТА LTE	118
А. Джинджиристий, М. Паламар АНАЛІЗ МЕТОДІВ КОЛАБОРАТИВНОЇ ФІЛЬТРАЦІЇ ПРИ ПІДБОРІ КОМАНДИ РОЗРОБНИКІВ КОМП'ЮТЕРНИХ СИСТЕМ	119
П. Євтух, В. Храпа МЕТОДИ ТА ЗАСОБИ УЗГОДЖЕННЯ ЦИФРОВИХ ПОТОКІВ ПРИ ЗАСТОСУВАННІ ІОТ ДО МЕДИЧНИХ ПРИЛАДІВ	120
А. Жуйвода МЕТОД ЗАБЕЗПЕЧЕННЯ ЯКОСТІ ПЕРЕДАЧІ ГОЛОСОВИХ СИГНАЛІВ В КОМП'ЮТЕРНИХ МЕРЕЖАХ	121
А. Жуйвода ПІДВИЩЕННЯ ЯКОСТІ ПЕРЕДАЧІ ПОТОКОВОГО ТРАФІКУ	122
О. Зимницький ВРАЗЛИВОСТІ РЕАЛІЗАЦІЇ КРИПТОГРАФІЧНИХ МЕТОДІВ ЗАХИСТУ ПРОТОКОЛУ SSL/TLS	123
Б. Калиниченко, І. Грод ДОСЛІДЖЕННЯ ВРАЗЛИВОСТЕЙ МЕРЕЖІ ОФІСУ "ZoomSupport" ТА МЕТОДІВ ЇХ УСУНЕННЯ	124
В. Ковальов, С. Лупенко МЕТОДИ ТА ЗАСОБИ ПОБУДОВИ КОМП'ЮТЕРИЗОВАНИХ ДІАЛОГОВИХ СИСТЕМ ТОРГОВОГО ЦЕНТРУ	125
І. Купратий НЕЙРОМЕРЕЖІ У СИСТЕМАХ БІОМЕТРИЧНОЇ АУТЕНТИФІКАЦІЇ ЗА КЛАВІАТУРНИМ ПОЧЕРКОМ	126
О. Ліщук, Є. Тиш ПЕРЕВАГИ ВИКОРИСТАННЯ ТЕХНОЛОГІЇ АГРЕГАЦІЇ КАНАЛІВ КОМП'ЮТЕРНОЇ МЕРЕЖІ	127

УДК 004.89

Р. Гаван, В. Яцишин

Тернопільський національний технічний університет імені Івана Пулюя

ОСОБЛИВОСТІ АНАЛІЗУ ТОНАЛЬНОСТІ ТЕКСТУ НА ОСНОВІ АСПЕКТІВ ПРИ ВДОСКОНАЛЕННІ КОМП'ЮТЕРНИХ СИСТЕМ

UDC 004.89

R. Havan, V. Yatsyshyn

(Ternopil Ivan Puluj National Technical University, Ukraine)

FEATURES OF SENTIMENT ASPECT-BASED ANALYSIS IN THE PROCESS OF COMPUTER SYSTEMS MODERNIZATION

Задача аналізу текстової інформації може розв'язуватись багатьма способами із застосуванням різних методів і підходів. Для оцінювання тональності текстових повідомлень у процесі вдосконалення комп'ютерних систем найбільш дієвим, з точки зору точності класифікації щодо приналежності відгуків до позитивного, негативного чи нейтрального класів, є аспектний підхід. Постановка задачі визначення тональності тексту за аспектами передбачає, що текст містить у собі оцінку деякого об'єкту. При цьому об'єкт розглядається як композиція його складових чи атрибутів, які можуть бути оцінені автором.

Зазвичай, враження про використання комп'ютерних систем, представляються у тематичних блогах або соціальних мережах, де основною одиницею тексту є відгуки користувачів, тобто текст є невеликим за об'ємом. У цьому випадку ефективним є використання підходу аспектного оцінювання думок. Аспекти можуть групуватися у категорії, для прикладу, інтерфейс користувача комп'ютерної системи, функціональність системи, поведінка системи тощо. Окрім цього, можлива наявність відгуків, де комп'ютерна система описується як єдиний об'єкт. Важливим при визначенні тональності текстової інформації є аналіз слів та виразів, які описують аспект сутності, так звані аспектні терміни.

Отже, задача аналізу текстової інформації у процесі вдосконалення комп'ютерних систем може бути поділена на ряд підзадач:

- визначення аспектних термінів;
- класифікація визначених термінів за аспектними категоріями;
- автоматичне визначення приналежності аспектів за категоріями.

Для виявлення аспектних термінів для конкретної предметної області, де використовується комп'ютерна система, можливе на основі декількох підходів.

Найбільш часто використовуваним типом аспектних термінів є явно задані терміни, які описують конкретний об'єкт, його частини або властивості, наприклад, візуальний дизайн, конкретну функцію системи, зручність використання системи та ін. Такі терміни найбільш часто виражаються іменниками або групами іменників, однак у деяких випадках, можливий опис явних аспектів за допомогою дієслів.

Іншим типом аспектних термінів є так звані неявні терміни, які представляються словами з явно вираженими оцінками компонента, які вказують на об'єкт відгуку. Зазвичай вони представляються прикметниками, наприклад, зручний (позитивний+ інтерфейс), швидкий (позитивний+відгук системи). Неявні аспектні терміни можуть комбінуватись з так званими операторами порівняння, які або змінюють або підсилюють значення оцінки щодо об'єкта або його властивості. Зокрема, це стосується застосування частки «не» (не дуже зручний, не надто швидкий). Важливість таких аспектних термінів для словників автоматичних систем аналізу тональності текстової інформації полягає в тому, що у ситуаціях неможливості розпізнавання термінів, що можуть бути пов'язані з помилками при наборі тексту, вживанні специфічних перевернутих термінів та ін. факторів неявні аспектні терміни дають змогу визначити ставлення автора по відношенню до деякої аспектної категорії.