Науковий вісник НЛТУ України
Scientific Bulletin of UNFU
ISSN 1994-7836 (print)
ISSN 2519-2477 (online)
https://doi.org/10.36930/40360316
Article received: 10.04.2026
Article accepted: 29.05.2026
Article published: 29.06.2026
Correspondence author: P. P. Maslianko masliankop@gmail.com

UDC 004.[02+89+942]

П. П. Маслянко, Ю. С. Мисак

Національний технічний університет України "Київський політехнічний інститут ім. Ігоря Сікорського", м. Київ, Україна

КОНЦЕПТУАЛЬНА МОДЕЛЬ NLP-СИСТЕМИ ТА NLP-ЗАСТОСУНОК СЕНТИМЕНТ-АНАЛІЗУ ВІДГУКІВ КОРИСТУВАЧІВ ПОСЛУГ ОПЕРАТОРІВ МОБІЛЬНОГО ЗВ'ЯЗКУ В УКРАЇНІ

Завдання та задачі з оброблення природної мови спрямовані на вирішення широкого класу наукових та прикладних завдань. З'ясовано, що систематизація таких завдань окреслює конкретні напрями проведення пошукових етапів дослідження, теоретичних та прикладних розробок в області NLP (англ. *Natural Language Processing*), насамперед таких, що найбільше обговорюють і дискутують в інженерній спільноті та R&D (англ. *Research and Development*) компанії. Зокрема – сегментація тексту і його морфологічний аналіз, синтаксичний і семантичний аналіз, машинний переклад і генерування тексту, сентимент-аналіз (або аналіз тональності тексту) і вибірковий пошук та аналіз інформації, а також багато інших прикладних застосувань. Встановлено, що актуальною є проблема дослідження та розроблення певних аксіоматичних інструментів подання NLP-систем на основі системного підходу та системної інженерії. Результат такого подання – науково обґрунтована концептуальна модель NLP-системи. Для розроблення концептуальної моделі NLP-системи було застосовано метод системної інженерії інформаційно-комунікаційних систем на основі бізнес-профілю Еріксона-Пенкера. Цей метод дає змогу формалізувати NLP-систему у вигляді множини конкретних іменованих класів сутностей, визначити проблеми, задачі і завдання NLP-системи, мету, ресурси, процеси і бізнес-правила, на основі яких така NLP-система функціонує. Така концептуальна модель NLP-системи упорядковує і систематизує потрібну і достатню множину класів сутностей і відношення між ними. Концептуальна модель NLP-системи дає змогу розробити структурне та динамічне подання моделі конкретної NLP-системи, показати внутрішню структуру її компонентів та інтерфейсів взаємодії між ними для вирішення конкретного класу завдань. Показано приклад застосування концептуального моделювання для розроблення концептуальної моделі NLP-системи та застосунку сентимент-аналізу (англ. *Sentiment Analysis*) відгуків користувачів послуг операторів мобільного зв'язку в Україні, проведено верифікацію та валідацію роботи застосунку. Перспективи подальших етапів дослідження спрямовані на вдосконалення концептуальної моделі NLP-систем та розроблення методик системної інженерії NLP-систем для вирішення конкретних класів завдань оброблення та аналізу природної мови.

Ключові слова: оброблення природної мови; бізнес-профіль Еріксона-Пенкера; системна інженерія NLP-систем; подання NLP-систем; концептуальна модель NLP-системи.

Вступ / Introduction

Серед багатьох напрямів застосування методів, моделей і засобів штучного інтелекту чільне місце належить методам і засобам оброблення природної мови. Оброблення природної мови NLP (англ. *Natural Language Processing*) – це діяльність, спрямована на її аналіз і синтез із застосуванням інформатики, штучного інтелекту, математичної та комп'ютерної лінгвістики та інших дисциплін. Таке визначення процедури оброблення природної мови наразі є надзвичайно поширеним у мережі Інтернет і часто трапляється в україномовних публікаціях, освітніх матеріалах та навчальних курсах з ІТ та штучного інтелекту.

Діяльність з оброблення природної мови спрямована на вирішення широкого класу наукових та прикладних завдань. Систематизація таких завдань окреслює конкретні напрями проведення пошукових етапів дослідження, теоретичних та прикладних розробок в області NLP, зокрема таких, що найбільше обговорюють і дискутують насамперед інженерна та університетська спільнота і R&D (англ. *Research and Development*) компанії.

1. Сегментація тексту [22]: поділ на речення – визначення меж між реченнями у тексті; токенізація – розбиття тексту на окремі слова або токени.
2. Морфологічний аналіз [17]: лематизація – зведення слова до його базової форми (леми); визначення частини мови – *в*изначення граматичної категорії слова (іменник, дієслово, прикметник і т. ін.).
3. Синтаксичний аналіз [46]: визначення залежностей – *а*наліз структури речення для визначення взаємозв'яз-

Маслянко Павло Павлович, канд. техн. наук, ст. наук. співробітник, доцент, кафедра прикладної математики.
Email: masliankop@gmail.com; https://orcid.org/0000-0003-4001-7811
Мисак Юрій Сергійович, магістр, кафедра прикладної математики.
Email: ygagagarin@gmail.com; https://orcid.org/0009-0008-2628-7982

Цитування за ДСТУ: Маслянко П. П., Мисак Ю. С. Концептуальна модель NLP-системи та NLP-застосунок сентимент-аналізу відгуків користувачів послуг операторів мобільного зв'язку в Україні. Науковий вісник НЛТУ України. 2026, т. 36, № 3. С. 146–160.
Citation APA: Maslianko, P. P., & Mysak, Yu. S. (2026). Conceptual model of NLP system and NLP application sentiment analysis of user feedback of mobile operators services in Ukraine. *Scientific Bulletin of UNFU*, 36(3), 146–160. https://doi.org/10.36930/40360316

ків між словами.

4. Семантичний аналіз [12]: визначення семантики – *р*озуміння значення слова або фрази у контексті; аналіз семантичної ролі – *в*изначення ролі слів у реченні (наприклад, підмет, присудок).
5. Розпізнавання іменованих сутностей (NER) [19, 25] – визначення та класифікація іменованих об'єктів, таких як особи, місця, організації, дати тощо.
6. Машинний переклад [47] – перетворення тексту з однієї мови на іншу.
7. Розпізнавання мовлення [28, 44] – перетворення усного мовлення на текстову форму.
8. Генерування тексту [23, 50] – створення природного мовлення або відповідей на певні запитання.
9. Сентимент-аналіз [29] – визначення та класифікація емоційного настрою тексту (позитивний, негативний, нейтральний).
10. Вибірковий пошук та аналіз інформації [26] – видобування інформації з тексту для подальшого використання.
11. Розрізнення схожих слів WSD (англ. *Word Sense Disambiguation*) [10] – визначення правильного значення слова в контексті, якщо воно має кілька значень.
12. Робота з діалоговими системами [15, 37] – розроблення систем, які можуть взаємодіяти з користувачами через природну мову.
13. Виявлення різних форм плагіату [39] – виявлення схожості між текстами для визначення можливого плагіату.

Зазначимо, що це перелік тільки основних, найпоширеніших завдань та задач у галузі NLP, і цей перелік постійно розширюється, додаючи нові бізнесові завдання та задачі з використанням нових технологій та підходів.

Далі, для прикладу проведення пошукових етапів дослідження та практичного застосування методів системної інженерії NLP-систем, було обрано завдання дослідження та розроблення концептуальної моделі NLP-системи та NLP-застосунку сентимент-аналізу (англ. *Sentiment Analysis*) відгуків користувачів послуг операторів мобільного зв'язку на множині даних коментарів у соціальних мережах за умов бізнес-правил, які обмежують поле даних на області операторів мобільного зв'язку в Україні.

Концептуальна модель NLP-системи та модель NLP-застосунку сентимент-аналізу відгуків користувачів базується на методології системної інженерії систем Data Science.

Об'єкт дослідження – розроблення концептуальної моделі NLP-системи та NLP-застосунку сентимент-аналізу відгуків користувачів послуг операторів мобільного зв'язку в Україні.

Предмет дослідження – методи і засоби розроблення концептуальної моделі NLP-системи та NLP-застосунку сентимент-аналізу відгуків користувачів послуг операторів мобільного зв'язку, які дадуть змогу інтегрувати множини класів сутностей, потрібних і достатніх для подання NLP-системи на концептуальному рівні.

Мета роботи – розробити концептуальну модель NLP-системи для формалізації та подання NLP-застосунків з оброблення природної мови, що дасть змогу отримати вербальну модель потрібної множини сутностей і процесів їх взаємодії задля виявлення недоліків й уточнення предмету дослідження.

Для досягнення зазначеної мети визначено такі основні *завдання дослідження*:

- формалізувати постановку завдання системної інженерії концептуальної моделі NLP-систем, визначити об'єкт, предмет, мету та остаточний результат дослідження, що дає можливість отримати вербальну модель потрібної множини сутностей і процесів їх взаємодії для реалізації теми статті;
- проаналізувати наявні практичні рішення і теоретичні інструменти інженерії NLP-систем, що дасть можливість дослідити стан ринку розробок та систематизувати результати останніх розробок інженерії NLP-систем задля виявлення недоліків й уточнення предмету дослідження;
- розробити та обґрунтувати концептуальну модель NLP-системи для сентимент-аналізу на основі бізнес-профілю Еріксона-Пенкера, що дає змогу інтегрувати множини класів сутностей, потрібних і достатніх для подання NLP-системи на концептуальному рівні;
- на основі концептуальної моделі NLP-системи розробити NLP-застосунок сентимент-аналізу відгуків користувачів послуг операторів мобільного зв'язку в Україні, що демонструє валідність запропонованої концептуальної моделі NLP-системи;
- провести верифікацію та валідацію NLP-застосунку, встановити обмеження застосування NLP-застосунку та особливості його роботи, що дає змогу продемонструвати практичну цінність отриманих теоретичних і практичних результатів.

Аналіз останніх досліджень та публікацій. Автори аналітичних звітів Brand24 [4] проаналізували результати моніторингу репутації в інтернеті та соціального прослуховування (англ. *social listening*) у всіх видах онлайн медіа і встановили, що його головна сильна сторона – можливість оперативного реагування на негативні відгуки та створення стратегій підвищення лояльності клієнтів на основі емоційних відтінків повідомлень.

Автори аналітичних досліджень результатів роботи MonkeyLearn [36], AI-платформи, що спеціалізується на управлінні досвідом клієнтів, співробітників компаній та громадян, встановили, що MonkeyLearn можна інтегрувати з різними додатками та сервісами для автоматизованого збирання та аналізу даних, що робить його універсальним, поширеним та зручним інструментом оброблення природної мови. Система пропонує користувачам як готові до використання моделі, так і можливості конструювання власних моделей без необхідності глибоких знань у галузі машинного навчання.

У книзі [20] автор проаналізував особливості застосування ChatGPT для інтелектуального оброблення текстів, в освіті та науці, мистецтві та повсякденному житті, програмуванні, використанні просунутими користувачами і застерігає фахівців від надмірних очікувань та встановлених обмежень застосування. Такі висновки мають забезпечити реальне розуміння можливостей та обмежень цих потужних технологічних інструментів.

У дослідженні [35] компанія SoftServe презентувала дослідження "Redefining the Future of Software Engineering", яке підготувала разом із журналом MIT Technology Review. Зокрема, тут показано, як агентний ШІ змінює підходи до розроблення програмних систем і стає інструментом розроблення AI-native інженерів. Прогнозується зростання застосування ШІ-агентів на всіх стадіях життєвого циклу до 72 %, зростання вимог до кваліфікації та кількості AI-інженерів до 51 %, зростання повної вартості володіння результатами розробок.

У роботі [27] автори досліджують проблему доопрацювання та налаштування NLP-моделей для виконання конкретного класу завдань оброблення природної мови, що потребує істотних додаткових витрат і ресурсів для розроблення NLP-систем і застосунків. Для вирішення цієї проблеми автори пропонують фреймворк CoDev, головна ідея якого полягає у колаборації та інженерії

локальної моделі для кожної концепції та глобальної моделі для інтеграції вихідних даних з усіма концепціями. Така ієрархічна колаборація моделей у багатьох випадках істотно спрощує розроблення та налаштування NLP-систем.

У роботі [21] автори досліджують проблему концептуального моделювання та проєктування інформаційних систем і застосунків, що потребує залучення стейкхолдерів і фахівців з різних предметних областей та інженерів знань, дотичних до призначення та області застосування систем або застосунків.

Автори роботи [32] встановили, що концептуальне моделювання інформаційних систем та застосунків для оброблення природної мови використовується, щоб допомогти експертам у предметній області та інженерам знань у виявленні вимог та побудові концептуальних моделей з тексту природною мовою. Автори визначають основну мотивацію, архітектуру, типи використовуваної природної мови (наприклад, обмежена та необмежена), типи згенерованої концептуальної моделі, підтримку перевірки специфікацій вимог та моделі подання знань. Автори також вказують на те, що наведені ними фреймворки не мають функції автоматичної верифікації результатів роботи.

Підводячи підсумок аналізу останніх досліджень та публікацій, можна зробити висновок про те, що наразі, на ринку починає домінувати AI-native підхід до розроблення NLP-систем. AI-native підхід до розроблення систем оброблення природної мови – це методологія, де штучний інтелект є не просто додатковою функцією, а основним ядром архітектури систем та застосунків з оброблення природної мови. На відміну від традиційного підходу, де ШІ інтегрується в готові системи як зовнішній, окремий інструмент, AI-native підхід розширює можливості інженерії NLP-систем та застосунків на основі:

1. Глибокої інтеграції (AI-First), коли система будується навколо можливостей нейромереж та великих мовних моделей (LLM), а не навколо класичної логіки чи жорстких алгоритмів.
2. Автоматизації процесів, коли моделі машинного навчання замінюють традиційні програмні ланцюжки (if-then), забезпечуючи гнучкість в інтерпретації та генеруванні мови.
3. Орієнтації на дані, коли пріоритет надається створенню інфраструктури для постійного навчання та адаптації системи на основі інтерфейсів взаємодії між сутностями системи.
4. Використання можливостей хмари, тобто аналогічно до "cloud-native", такий підхід передбачає використання масштабованої інфраструктури, що може бути оптимізована для високих навантажень ШІ-обчислень.
5. Інтерактивності та контекстності, тобто коли системи розробляються для розуміння складних нюансів, намірів та контексту завдань та запитів користувача, що робить взаємодію із системою максимально наближеною до людського спілкування.

Такий підхід дає змогу створювати NLP-рішення (чат-боти, аналітичні системи, перекладачі), які можна верифікувати і валідувати та адаптувати для вирішення конкретних завдань і задач користувачів. Зокрема, у сфері моніторингу соціальних медіа і сентимент-аналізу застосунки Brand24 і MonkeyLearn [4, 36] є визнаними лідерами на ринку, надаючи користувачам потужні інструменти для виявлення клієнтських настроїв. Однак, до недоліків цих потужних інструментів варто віднести істотну надлишкову функціональність і вимоги до потужності обчислювальних ресурсів, відчутну повну вартість володіння та обмеженість застосування для вирішення вузько-профільних завдань аналізу природної мови, зокрема української.

Матеріали та методи дослідження. У роботі використано Аксіоматичний метод наукового пізнання, системний підхід та методологію системної інженерії інформаційно-комунікаційних систем на основі бізнес-профілю Еріксона-Пенкера. Для сентимент-аналізу застосовано модель багатомовного подання XLM-R (англ. *Cross-lingual Language Model* – RoBERTa) та методи оброблення природної мови.

Результати дослідження та їх обговорення / Research results and their discussion

Покажемо основні аксіоматичні положення, припущення та поняття пошукових етапів дослідження що, з нашої точки зору, потрібні для наукового обґрунтування методів, моделей і засобів для розроблення концептуальної моделі NLP-систем. Тут і надалі будемо спиратись на ідеї та результати про поняття і сутності складових систем Data Science [32]. Для розроблення концептуальної моделі NLP-системи будемо застосовувати аксіоматичний метод наукового пізнання.

Поняття і сутності NLP-систем. Спираючись на загальне визначення сутностей моделі Data Science [11], перейдемо до визначення такої підобласті Data Science, як оброблення природної мови NLP (англ. *Natural Language Processing*).

Перш ніж заглиблюватися в особливості інженерії NLP-систем, потрібно визначитися з поняттям "система оброблення природної мови". На підставі Структурного подання Data Science у форматі діаграми Венна [32], на нашу думку, внаслідок аналізу предметної області NLP концепцію NLP можна розкласти на три класи сутностей:

1. Множина сутностей "Природна мова".
2. Множина сутностей "Інформатика, Штучний інтелект, математична та комп'ютерна лінгвістика".
3. Множина сутностей "Проблеми і завдання аналізу та синтезу природної мови".

Далі запропоновано авторське визначення класів сутностей подання предметної області концепції NLP-систем у вигляді діаграми Венна (рис. 1,*a*) та конкретизації концепції NLP-систем у вигляді NLP-систем/застосунків сентимент-аналізу (ССА) (див. рис. 1,*b*).

1. *Сутності множини N – природна мова.* Множина N – сукупність сутностей природної мови в межах концепції NLP визначається як множина N, де $n \in N$ охоплює всі наявні елементи та аспекти, що стосуються розпізнавання, аналізу та розуміння природної мови. Сутності цієї множини – це її окремі елементи, тобто самі числа. Опис природною мовою означає, що їх пояснюємо або називаємо ці числа звичайними людськими словами, а не математичними знаками чи формулами. До цієї множини входять лінгвістичні одиниці, такі як слова, фрази та речення, а також мовні структури, наприклад, синтаксичні та семантичні взаємозв'язки. До складу N входять також сутності, пов'язані із семантичним аналізом, такі як іменовані сутності, події, атрибути та інші лінгвістичні конструкції.

   Окрім цього, множина N враховує контекстуальні аспекти, такі як емоційний тон, стилістика та інші мовні нюанси, що допомагають досягти точнішого розуміння тексту та взаємодії з природною мовою в широкому спектрі застосувань, враховуючи розроблення систем автоматичного перекладу, аналізу відгуків користувачів та створення інтелектуальних асистентів.

2. *Сутності множини A – інформатика, штучний інтелект, математична та комп'ютерна лінгвістика.* Множина *A* – сукупність цих сутностей у рамках концепції NLP утворює множину *A*, де $a \in A$ і визначається як усі об'єкти, що входять до складу цих галузей науки. До множини *A* вносяться концепції та методи, розроблені в галузі інформатики для оброблення природної мови, враховуючи алгоритми машинного навчання, статистичні моделі та техніки оброблення природної мови. Ця множина охоплює інструментальні засоби та технології, що використовуються для оброблення природної мови, такі як алгоритми машинного навчання, моделі глибокого навчання, а також лексичні та семантичні ресурси. Також охоплює підходи, що базуються на принципах ШІ, зокрема, системи оброблення природної мови, які здатні адаптуватися та вдосконалювати свої навички залежно від нових даних. У контексті математичної та комп'ютерної лінгвістики, до множини A входять формальні моделі мови, граматичні структури та методи аналізу тексту, що сприяють розвитку та вдосконаленню систем оброблення природної мови для досягнення більшої точності та зрозуміння мовлення.

a) *b*)

Рис. 1. NLP-системи у вигляді діаграми Венна / NLP systems represented as Venn diagrams: *a*) концепт NLP-системи / Concept of NLP systems; *b*) концепт NLP-системи для сентимент-аналізу / Concept of NLP systems sentiment analysis

3. *Сутності множини P – проблеми і завдання аналізу та синтезу природної мови.* Множина *P* – у концепції NLP містить різноманітні проблеми, завдання та задачі, пов'язані із здатністю систем оброблення природної мови ефективно аналізувати та синтезувати мовні конструкції. До цієї множини входять важливі елементи, спрямовані на вирішення практичних завдань, пов'язаних з обробленням тексту та мовної інформації.

До складу множини *P* входять завдання розпізнавання мовленнєвих патернів, класифікації тексту, визначення семантики, а також завдання аналізу тональності та емоційного відтінку в мовленні. Завдання автоматичного перекладу, створення резюме текстів, аналізу сентименту та генерування природної мови також є невід'ємною частиною цієї множини.

Окрім цього, вона враховує високорівневі завдання, такі як розрізнення амбігвітету, робота з невизначеністю в тексті та інші аспекти, що роблять аналіз та синтез природної мови глибшим та повнішим.

На діаграмі Венна на рис. 1,*a* наведено перетини множин 1-2, 1-3 та 2-3 класів сутностей *N*, *A* та *P* утворюють інтегровані міждисциплінарні ресурси (сутності), виділені за ознакою наявності функціональних відношень між ресурсами (сутностями) різних типів, зокрема Set N, Set A, Set P.

***Інтегровані міждисциплінарні ресурси* (*1-2*) (*IMP 1-2*)** [32] – підмножина, яка утворюється внаслідок перетину множини *N* – природна мова з множиною *A* – інформатика, штучний інтелект (ШІ), математична та комп'ютерна лінгвістика і може бути охарактеризована як множина елементів, що мають властивості інтеграції та інтероперабельності. Нехай цю сутність буде позначено як множина NA, де NA∈ представляє концепції та аспекти, що є спільні для природної мови й інформатики, штучного інтелекту, математичної та комп'ютерної лінгвістики.

Такий перетин відображає взаємодію та взаємопроникнення областей, спрямованих на вдосконалення розуміння та оброблення мовленнєвої інформації за допомогою сучасних технологій та аналітичних підходів.

***Інтегровані міждисциплінарні ресурси* (*1-3*) (*IMP 1-3*)** [32] – підмножина, яка утворюється внаслідок перетину множини *N* – природна мова і множини *P* – проблеми і завдання аналізу та синтезу природної мови і визначається як множина елементів, що об'єднує концепції та виклики, що виникають у процесі розуміння, оброблення та створення мовного контенту. Нехай цю сутність буде позначено як множина NP, де NP∈ відображає інтерсекцію областей природної мови і завдань аналізу та синтезу мовленнєвої інформації.

До цієї множини входять завдання, такі як розпізнавання та аналіз семантики тексту, класифікація мовленнєвих патернів, визначення ентітетів та сутностей, а також синтез мовленнєвого висловлення. Множина *NP* охоплює проблеми визначення сентименту та емоційного відтінку в тексті, а також містить завдання автоматичного перекладу та генерування природної мови.

***Інтегровані міждисциплінарні ресурси* (*2-3*) (*IMP 2-3*)** [32] – сутність, що утворюється внаслідок перетину множини *A* – інформатика, штучний інтелект (ШІ), математична та комп'ютерна лінгвістика і множини *P* – проблеми і завдання аналізу та синтезу природної мови, визначається як множина елементів, які об'єднують концепції та завдання, що виникають за використання Інформатики, ШІ, Математичної та комп'ютерної лінгвістики для аналізу та синтезу природної мови. Нехай цю сутність буде позначено як Множина AP, де AP∈ ві-

дображає спільні елементи областей Інформатики, ШІ, Математичної лінгвістики і завдань аналізу та синтезу природної мови.

Множина АР містить аспекти розв'язання задач аналізу та синтезу природної мови з використанням інструментів і методів інформатики та ШІ. Це охоплює розроблення та використання алгоритмів машинного навчання, моделей глибокого навчання, оброблення природної мови з використанням математичних методів, а також комп'ютерно-лінгвістичні підходи до аналізу текстової інформації.

Підсумовуючи можна стверджувати, що було отримано достатнє, для поточної стадії розробки, наукове обґрунтування та концепції NLP-систем, і концепції NLP-систем сентимент-аналізу.

Концептуальна модель NLP-системи для сентимент-аналізу на основі бізнес-профілю Еріксона-Пенкера. Для розроблення концептуальної моделі NLP-системи було застосовано метод системної інженерії інформаційно-комунікаційних систем на основі бізнес-профілю Еріксона-Пенкера [11].

Цей метод дає змогу формалізувати NLP-систему у вигляді множини конкретних класів сутностей, визначити проблеми, мету виконання, ресурси, процеси і бізнес-правила, на основі яких така система функціонує [11, 31] (рис. 2).

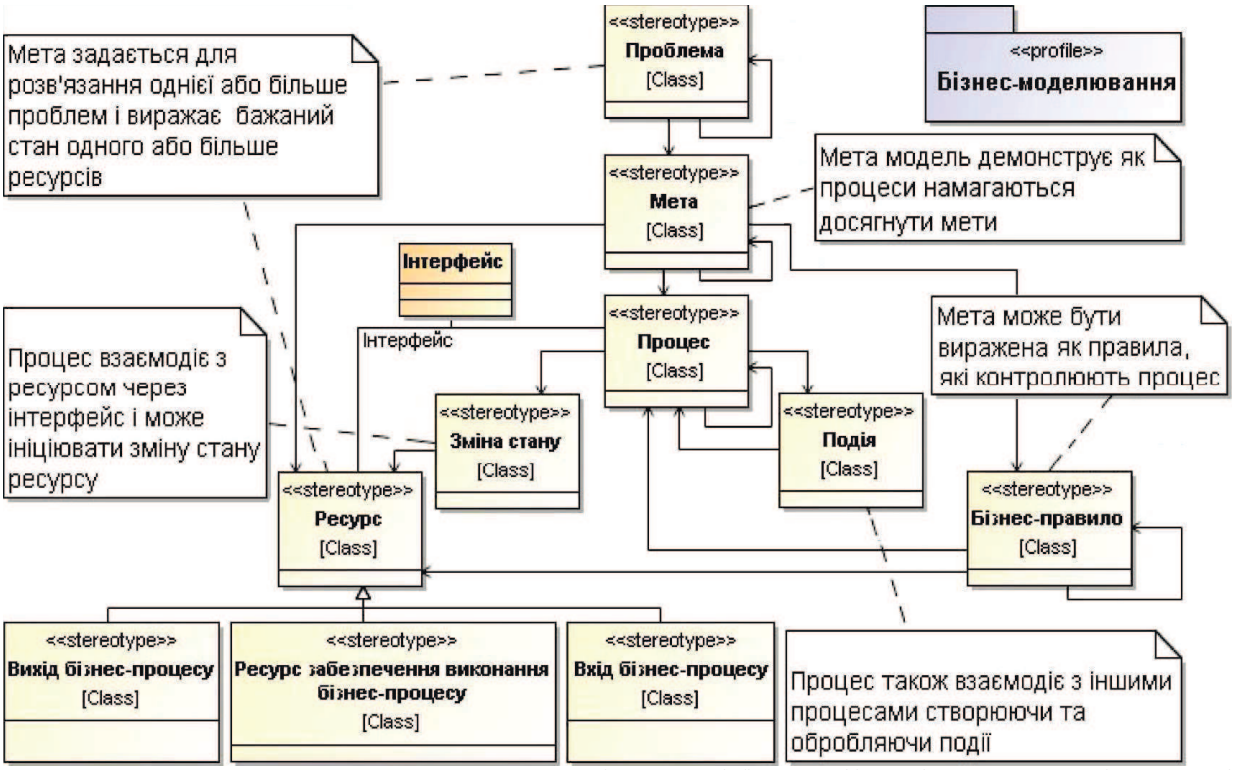

Рис. 2. Концептуальна модель NLP-системи для сентимент-аналізу на основі бізнес-профілю Еріксона-Пенкера. Діаграма класів у нотації UML / Conceptual model of an NLP system for sentiment analysis based on the Erickson-Penker business profile. Class diagram in UML notation [11, 31]

Формалізуємо кожен із класів концептуальної моделі NLP-системи для сентимент-аналізу.

Клас *Проблема* – актуальне питання, що потребує відповідних рішень; основна мотивація розроблення системи, яка спонукає до формулювання конкретної мети*: Необхідність наукового обґрунтування методів, моделей і засобів для розроблення концептуальної моделі NLP-систем.*

Клас *Мета* – виражає глобальну ціль роботи NLP-системи, покликана вирішити поставлену проблему: *Розроблення науково обґрунтованої концептуальної моделі NLP-системи для продукування NLP-застосунків сентимент-аналізу.*

Клас *Процес* – складові загальної діяльності системи, внаслідок якої досягається мета; чітко визначена послідовність дій/підпроцесів, що приводить до виконання певного завдання: *На підставі міжнародного стандарту Cross Industry Standard Process for Data Mining* (*CRISP-DM*)*, що є унікальною моделлю, яка слугує основою для оброблення даних.*

Клас *Зміна стану* – можливі зміни певних ресурсів унаслідок роботи певних процесів: *Відповідно до потреб процесів.*

Клас *Ресурс* – будь-які сутності (матеріальні чи нематеріальні), що споживаються, застосовуються та продукуються *NLP*-системою:

- клас *Вихід бізнес-процесу* – ресурси, що продукуються системою, остаточний результат її функціонування;
- клас *Ресурс забезпечення виконання бізнес-процесу* – ресурси, що забезпечують виконання процесів, але не є остаточним результатом її роботи: *Інструменти оброблення текстових даних, інструменти збереження й оновлення моделі ССА, інструменти класифікації*;
- клас *Вхід бізнес-процесу* – первинні ресурси входу початкових процесів, які ініціалізують цикл роботи системи: *вибір моделі сентимент-аналізу, сформовані множини, метрики множин, необроблені текстові дані, ініціалізована ССА.*

Клас *Подія* – виникає внаслідок певних зовнішніх факторів, або ж як результат взаємодії між процесами: зміна навчальних текстових даних, що впливає на *Навчання ССА*; поява нової найефективнішої версії *Навченої ССА* (*Класифікатор*) у процесі *Навчання ССА*, яка, як наслідок, замінить поточну версію, що бере участь у процесі *Сентимент-аналіз*; введення періоду і компанії для аналізу користувачем (запит користувача) у процесі

Візуалізація, що спонукає систему до їхнього зображення результатів;

Клас *Бізнес-правило* – формальні інструкції, що регулюють, обмежують і встановлюють контекст та рамки функціонування NLP-системи.

Модель класу "Процес" концептуальної моделі NLP-системи для сентимент-аналізу на основі бізнес-профілю Еріксона-Пенкера. Модель класу "Процес" концептуальної моделі NLP-системи для сентимент-аналізу (рис. 3) розроблено на підставі міжнародного стандарту CRISP-DM [13], також відомого як Cross Industry Standard Process for Data Mining, що є унікальною моделлю, яка слугує основою для оброблення відповідних даних.

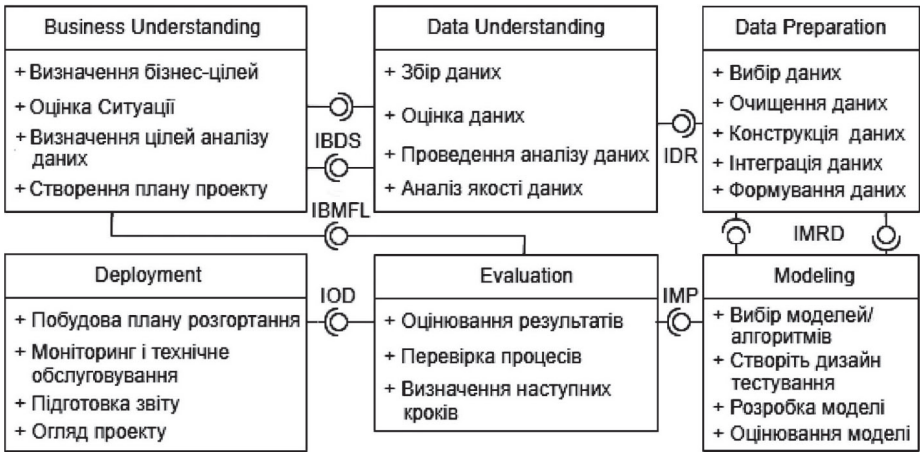

Рис. 3. Модель класу "Процес" концептуальної моделі NLP-системи для сентимент-аналізу на основі бізнес-профілю Еріксона-Пенкера. Діаграма класів у нотації UML / Class model "Process" of the conceptual model of the NLP for sentiment analysis system based on the Erickson-Penker business profile. Class diagram in UML notation

Надані і затребувані інтерфейси взаємодії між компонентами Моделі класу "Процес" забезпечують функціонування моделі відповідно до встановлених бізнес-правил.

Модель NLP-системи та NLP-застосунку сентимент-аналізу відгуків користувачів послуг операторів мобільного зв'язку в Україні. Модель NLP-системи наведено у вигляді структурного та динамічного подань, що характеризують NLP-систему з різних точок зору.

Модель NLP-системи. Структурне подання. На рис. 4 подано модель NLP-системи для сентимент-аналізу, що складається з п'яти компонентів.

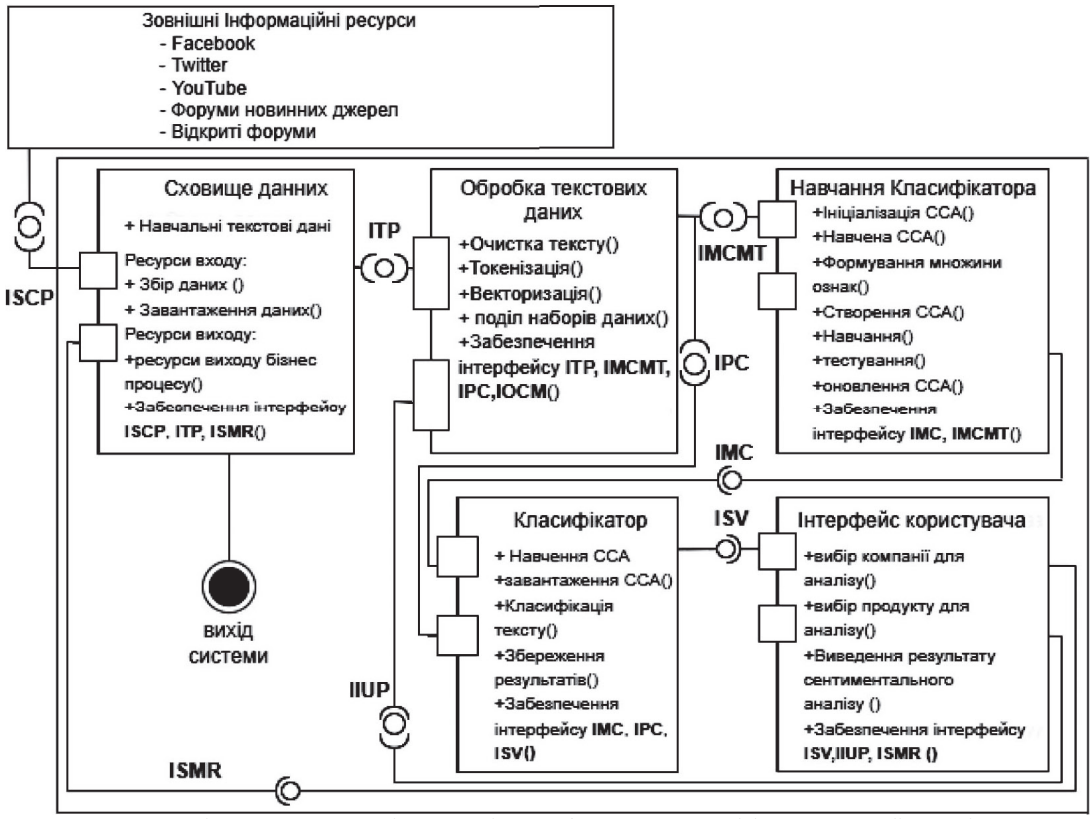

Рис. 4. Модель NLP-системи. Діаграма компонентів у нотації UML / NLP system model. Component diagram in UML notation

Компонент "Сховище даних/Озеро даних": Компонент "Сховище даних/Озеро даних" в інформаційній системі призначений для зберігання та отримання потрібних текстових даних зі зовнішніх ресурсів для подальшого оброблення текстових даних компонентом "Оброблення текстових даних" та зберігання результатів роботи моделі сентимент-аналізу. Тут варто пояснити, що сховище даних DWH (англ. *Data Warehouse*) та озеро даних DL (англ. *Data Lake*) – це системи для зберігання великих обсягів даних, але з різною метою:

DWH зберігає структуровану, готову до аналізу інформацію, а озеро даних – сирі, неструктуровані дані у власному форматі. Озеро даних забезпечує гнучкість доступу до даних, а сховище – їхню точність для бізнес-звітів.

Компонент "Оброблення текстових даних" відповідає за оброблення та підготовку текстових даних, які отримуються від компонента "Сховище даних/Озеро даних" для подальшого навчання класифікатора або використання у завданнях NLP. Компонент може бути реалізованим на основі інструментів інформатики, штучного інтелекту, математичної та комп'ютерної лінгвістики. Основними функціями цього компонента є: попередній аналіз даних, очищення та нормалізація даних, токенізація, стемінг, вилучення ключових слів, видалення стоп-слів, векторизація даних.

Компонент "Навчання класифікатора" є ключовою частиною системи для завдань сентимент-аналізу текстів, де мета полягає в тому, щоб навчити модель класифікувати текстові дані на позитивний, негативний або нейтральний сентименти.

Компонент "Класифікатор" відповідає за класифікацію сентименту в текстах на позитивний, негативний або нейтральний. Цей компонент використовує останню версію моделі, яку було навчено в компоненті "Навчання класифікатора", на підготовлених даних, що пройшли оброблення в компоненті "Оброблення текстових даних".

Компонент "Інтерфейс користувача" надає доступ до результатів аналізу сентименту текстів, які були оброблені й класифіковані компонентом "Класифікатор". Для забезпечення змістовного подання результатів у компоненті забезпечуються елементи керування та візуалізації. Наприклад, елемент вибору параметрів для аналізу (вибір компанії, послуги, моделі сентимент-аналізу), графічні елементи візуалізації результатів (графіки часового ряду або агреговані значення у вигляді секторної діаграми), елементи сортування та фільтрації та ін.

Інтерфейси NLP-системи. Взаємодія компонентів NLP-системи забезпечується за допомогою восьми затребуваних і наданих інтерфейсів:

- ISCP (англ. *Interface Source Connection Protocol*) – інтерфейс виконання протоколу під'єднання до зовнішніх інформаційних ресурсів, а також надавання функціоналу отримання текстових даних на вхід процесу "Сховище даних/Озеро даних";
- ITP (англ. *Interface Text Processing*) – інтерфейс передавання даних для сентимент-аналізу, а також завантажених в оперативну пам'ять текстових даних на вхід процесу текстового оброблення за вказаними параметрами користувача;
- IMCMT (англ. *Interface Machine Classification Model Training*) – інтерфейс передавання вихідних (із модуля текстового оброблення) опрацьованих (очищених, нормалізованих, токенізованих, нумеризованих) навчальних текстів на вхід процесу навчання системи для сентимент-аналізу;
- IPC (англ. *Interface Processing Classification*) – інтерфейс передавання вихідних (із модуля текстового оброблення) опрацьованих (очищених, нормалізованих, токенізованих, анотованих) текстових даних, введених користувачем фільтра, на вхід модуля безпосереднього Класифікатора;
- IMC (англ. *Interface Machine Classification*) – інтерфейс завантаження останньої найефективнішої версії збереженої навченої ССА для її подальшого використання у процесі Класифікатор попередньо обробленого тексту, враховуючи параметри користувача;
- IIUP (англ. *Interface Input User Parameters*) – інтерфейс передавання результату обрання користувачем інших параметрів для сентимент-аналізу тексту від Інтерфейсу користувача до оброблення текстових даних;
- ISV (англ. *Interface Solve Visualization*) – інтерфейс передавання результатів сентимент-аналізу тексту від Класифікатора до Інтерфейсу користувача;
- ISMR (англ. *Interface Save Modeling Result*) – інтерфейс передавання вихідних ресурсів проведеного сентимент-аналізу на зберігання компоненту сховища даних.

Модель NLP-системи. Динамічне подання. Динамічне подання NLP-системи для сентимент-аналізу показано у вигляді діаграми діяльності (рис. 5) відповідно до функціональності компонентної моделі (див. рис. 4). Така модель NLP-системи відображає як внутрішню структуру кожного компонента, так й інтефейси взаємодії компонентів. Діаграма діяльності з "водними доріжками" є важливою частиною моделі Еріксона-Пенкера, оскільки вона поєднує обидва види подань та формалізує діяльність NLP-системи.

Динамічне подання моделі NLP-системи регламентується і діаграмами взаємодії, зокрема діаграмою послідовності, діаграмою часу та діаграмою зв'язку/співпраці. Модель NLP-системи для сентимент-аналізу у вигляді діаграми послідовності показано на рис. 6. Діаграма послідовності для NLP-системи сентимент-аналізу описує покроковий шлях від моменту, коли користувач вводить текст, до отримання результату оцінки емоційного забарвлення (позитивний, негативний чи нейтральний). На цьому рисунку наведено модель такої взаємодії у вигляді графічного коду, який автоматично візуалізується у багатьох сучасних редакторах, а також детальний опис кожного кроку.

Так було отримано модель NLP-системи для сентимент-аналізу, що складається з потрібної кількості подань і відповідних видів математичного, інформаційного, програмного та інших необхідних видів забезпечення внутрішньої структури компонентів. Далі, на основі такої моделі, можна розробляти необхідні NLP-застосунки для вирішення конкретних завдань сентимент-аналізу.

NLP-застосунок сентимент-аналізу відгуків користувачів послуг операторів мобільного зв'язку в Україні. Покажемо особливості імплементації моделі NLP-системи у вигляді NLP-застосунку сентимент-аналізу відгуків користувачів послуг операторів мобільного зв'язку в Україні. Оскільки підходи до розроблення NLP-систем сентимент-аналізу, де дані можуть містити тексти різними мовами (українська, англійська та ін.), вносять додаткові виклики, пов'язані з обробленням багатомовності, то потрібно додатково дослідити й інженерні рішення/застосунки. Зокрема, наразі існують кілька типових можливих рішень такої NLP-системи:

1. *Багатомовне подання.* Використання моделей, таких як mBERT (Multilingual BERT) [6, 42] або XLM-R [2, 8], які можуть кодувати тексти з різних мов у спільний багатомовний простір. Ці моделі тренуються на великих багатомовних датасетах і здатні розпізнавати й узагальнювати міжмовні регулярності та взаємозв'язки.
2. *Підтримка багатомовного оброблення.* Потрібно використовувати окремі NLP-моделі для кожної мови.
3. *Аналіз з Машинним перекладом.* Ще одним варіантом є використання сервісів машинного перекладу для перекладення всіх текстів на одну "основну" мову, якою у вас уже розроблений або доступний надійний сентимент-аналіз. Цей підхід має ризик втрати нюансів сентименту через неточності в перекладі.

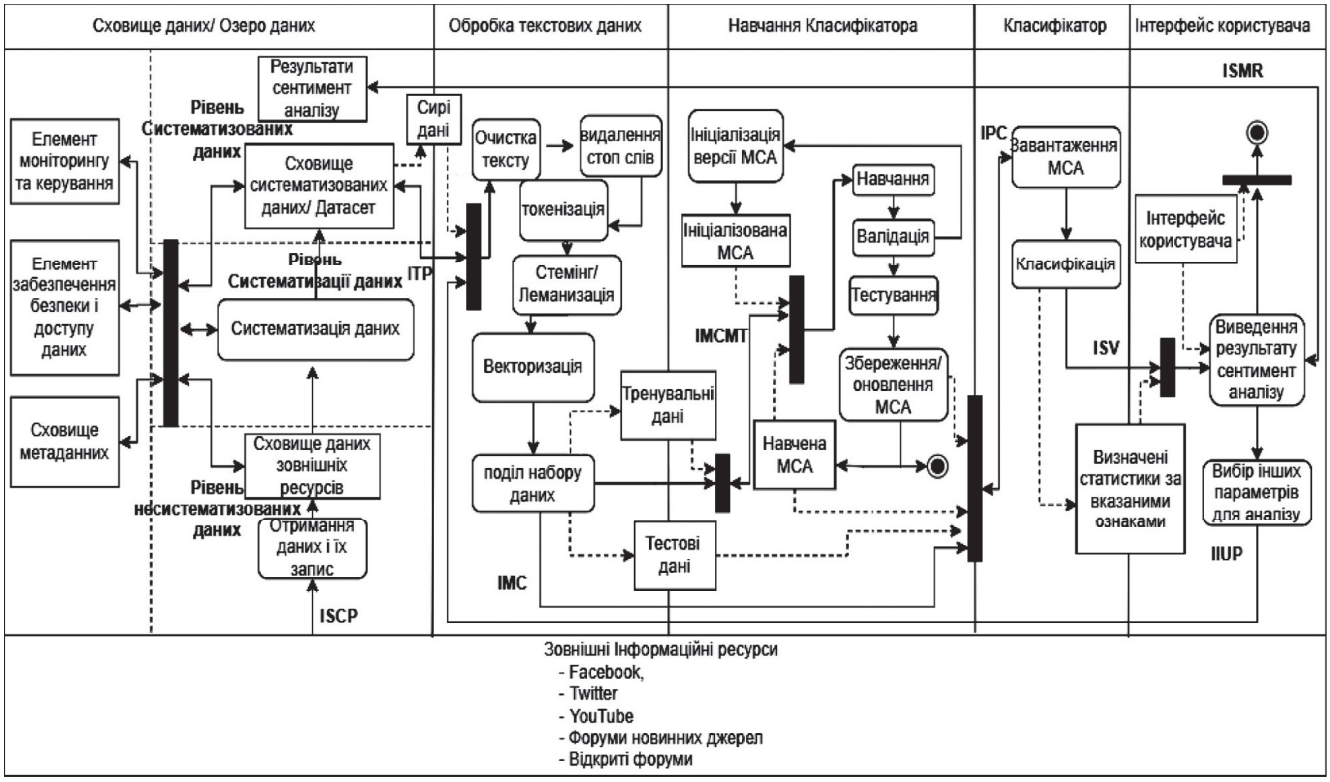

Рис. 5. Модель NLP-системи. Діаграма діяльності в нотації UML / NLP system model. Activity diagram in UML notation

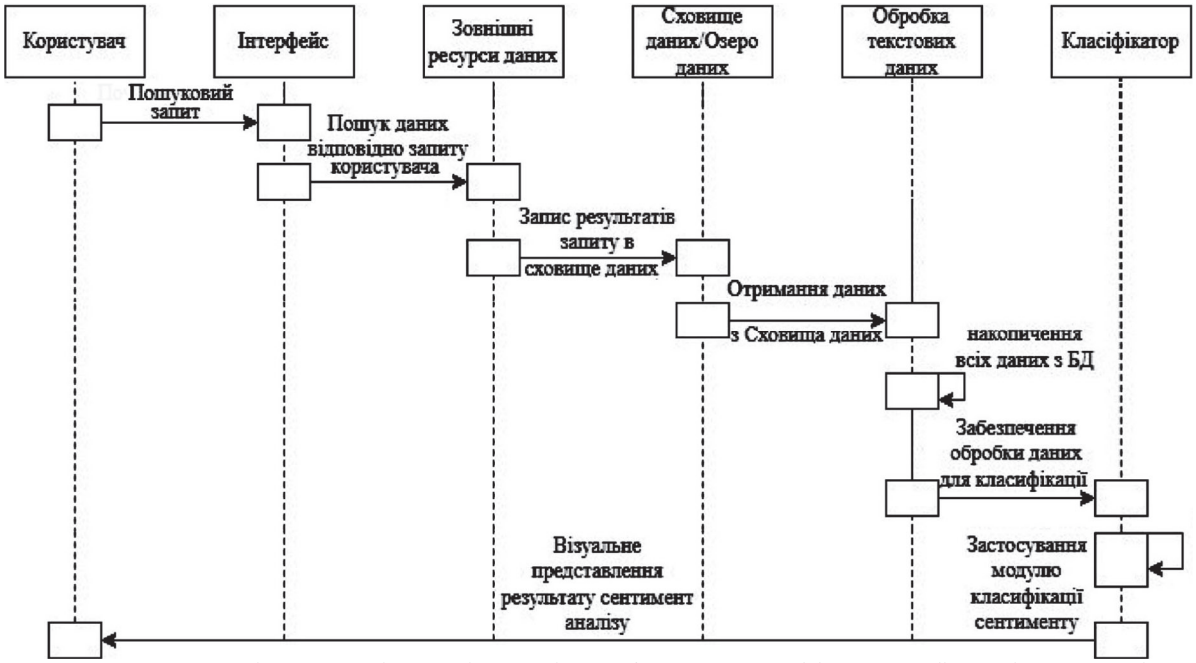

Рис. 6. Модель NLP-системи. Діаграма послідовності у нотації UML / NLP system model. Sequence diagram in UML notation

Було запропоновано реалізацію таких систем на принципах Трансферного навчання. Трансферне навчання [40] (англ. *Transfer Learning*) є методологією в машинному навчанні, яка передбачає використання знань, набутих під час вирішення одного завдання для розв'язання іншого, але схожого завдання.

Короткий експрес-аналіз (таблиця) показав, що значно краще, з даними саме нашого завдання, справляється модель на основі багатомовного подання. Експрес-аналіз варіантів рішень або застосунків – це швидке порівняння програм за головними критеріями для вибору найкращого варіанта. Такий аналіз допомагає аналітику знайти ідеальний інструмент без зайвих витрат часу та грошей, який допоможе йому швидко оцінити будь-які програми чи сервіси.

Таблиця. Експрес-аналіз варіантів рішень/застосунків / Express analysis of solution/application options

№	Варіант рішення/застосунку	Метрики точності
1	Багатомовне подання системи	Accuracy =82,8 %; Precision = 83,6 %; Recall =86,3 %; F1 = 83,8 %;
2	Підтримка багатомовного оброблення	Accuracy =58,6 %; Precision = 68,6 %; Recall =61,0 %; F1 = 55,8 %;
3	Аналіз з Машинним перекладом	Accuracy =54,5 %; Precision = 61,6 %; Recall =61,6 %; F1 = 54,7 %;

Опис даних. Результати експрес-аналізу було розраховано на основі даних, отриманих із соціальної мережі Facebook за допомогою налагодженого отримання даних через API Facebook [33]. Дані містять коментарі ко-

ристувачів соціальної мережі, які коментували дописи операторів мобільного зв'язку в Україні, і датуються за період з 20.11.2021 по 24.02. 2023 роки. Цей період охоплює коментарі як до, так і після повномасштабного вторгнення. Загальна кількість коментарів, які було отримано зі сторінок трьох головних операторів мобільного зв'язку в Країні, сягає понад 24 тис. коментарів.

Основні характеристики даних:

- post_id – унікальний ідентифікатор посту, під яким створено коментар;
- comment – текст коментаря користувача;
- Company_Tag – тег компанії, до якої належить коментар (ks – Київстар, vf – Vodafone, life – Lifecell);
- time – час створення коментаря.

post_id	comment	time	Com
5,89487E+15	дякую за вашу роботу! разом-до пер	2023-02-2	life
5,89487E+15	viktor viktor	2023-02-2	life
5,89487E+15	ви найкращі!	2023-02-2	life
5,89487E+15	і чому не можна поговорити з операт	2023-02-2	life
5,89487E+15	сергей кифорук	2023-02-2	life
5,89487E+15	звьязок погіршився. багато де перест	2023-02-2	life
5,89487E+15	доброго ранку! сьогодні перший ден	2023-02-2	life
5,89487E+15	дякую загарний звязок гарной робот	2023-02-2	life
5,89487E+15	як був поганий зв'язок таким і залиш	2023-02-2	life
	увага! через вплив економічних чинн		
5,89487E+15	рrс-рс	2023-02-2	life
5,89487E+15	цены на услуги сделали больше всех	2023-02-2	life
5,89487E+15	звязок поганий,лайф погано ловить,п	2023-02-2	life
5,89842E+15	євген поладич	2023-02-2	life
5,89842E+15	лайфсел как связаться с вашими опер	2023-02-2	life
5,89279E+15	традиційний вечірній 2g😠	2023-02-2	life
5,89279E+15	для цього потрібний як мінімум інтер	2023-02-2	life

Рис. 7. Фрагмент набору даних / A fragment of the dataset

Далі покажемо реалізацію математичного забезпечення NLP-застосунку сентимент-аналізу відгуків користувачів послуг операторів мобільного зв'язку в Україні на основі моделі багатомовного подання.

Математичне забезпечення та імплементація компонентів NLP-застосунку сентимент-аналізу відгуків користувачів послуг операторів мобільного зв'язку в Україні. Зазначимо, що для розроблення математичного забезпечення не можемо використати класичні методи класифікації наївний Баєсовий класифікатор [7], метод опорних векторів (SVM) [14] та інші, оскільки наші дані не мають міток сентименту і тому було використано методи, які базуються на навчанні на цільових значеннях. Було запроповано шляхи адаптації підходів під українську мову, а також методи попереднього оброблення текстів. Висвітлюємо структурування даних, процеси навчання та оцінювання алгоритмів, а також пропонуємо технічні аспекти реалізації NLP-системи, зокрема й обрання інструментарію та архітектури застосунку.

Компонент "Сховище даних/Озеро даних": у цьому компоненті було обрано тип архітектурного підходу "Озеро даних" (англ. *Data Lake*) [9, 18] до побудови процесу збирання даних із зовнішніх ресурсів, збереження неструктурованих даних, таких як коментарі користувачів соціальних мереж і результату роботи моделі сентимент-аналізу.

Архітектуру було обрано за такими критеріями, як: гнучкість, масштабованість, підтримка машинного навчання [41]. Для збирання текстових даних з будь-яких відкритих ресурсів використовуються відповідні APIs або завантаження наборів даних за допомогою власних систем збирання.

Запропоновано системи зберігання великих даних: розподілена файлова система – Hadoop Distributed File System (HDFS) [3], або представники хмарних систем Amazon S3 [1] і Azure Data Lake Storage [34]. Забезпечення безпеки та доступу: цей елемент відповідає за керування доступом до даних у Data Lake. Він забезпечує захист від несанкціонованого доступу, керує правами доступу, шифрує дані та забезпечує аудит безпеки.

Моніторинг та керування: цей елемент дає змогу відстежувати стан системи Data Lake, моніторити продуктивність, керувати обсягом даних та оптимізувати розподілені обчислення.

Зберігання метаданих: метадані, які описують структуру даних, схеми, джерела даних тощо, зберігаються в окремому репозиторії метаданих. Цей елемент допомагає керувати та описувати дані, збережені в Data Lake.

Компонент "Оброблення текстових даних" має завдання стандартизувати різноманітний набір текстових даних. Кожен початковий документ у вхідному корпусі оброблення текстових даних проходить через такі етапи стандартизації.

1. Очищення тексту.
2. Токенізація – використання токенізатора, специфічного для XLM-R, щоб розділити текст на субворди-токени.
3. Додавання спеціальних токенів – автоматичне додавання [CLS], [SEP] і [PAD] токенів за допомогою токенізатора, де [CLS] ставиться перед кожною послідовністю для завдань класифікації, [SEP] розділяє кожну пару текстів і [PAD] позначає місця подовження.
4. Оброблення довжини послідовності – встановлення єдиної довжини послідовності у спосіб обрізання чи додавання [PAD] токенів з використанням параметрів max_length і padding відповідно.
5. Створення спеціальної маски – генерування маски уваги (attention mask), яка вказує моделі ігнорувати [PAD] токени під час оброблення послідовностей.

Компонент "Навчання класифікатора". Дослідження варіантів моделей показало, що моделі багатомовного подання працюють на наших даних краще. Опишемо принцип роботи моделі XLM-R яка обрана як основна модель у цьому компоненті. Модель XLM-R (Cross-lingual Language Model – RoBERTa) [2, 8] – це сучасна модель оброблення природної мови (Natural Language Processing, NLP).

Основні властивості моделі:

Багатомовність. На відміну від багатьох ранніх моделей, що обмежувалися однією чи кількома мовами, XLM-R тренувалася на даних, отриманих з більш ніж ста мов, що робить її ефективною для багатомовного оброблення тексту.

Технологія трансформера. XLM-R використовує архітектуру "трансформера", сучасний підхід в області NLP, який опирається на так звані механізми уваги (attention mechanisms). Ці механізми дають змогу моделі зосередитися на різних частинах тексту та зв'язках між словами для глибшого розуміння контексту.

Масштабне тренування. Замість тренування з нуля, XLM-R використовує техніку попереднього навчання (pretraining) на великій кількості текстів з різних мов і джерел. Це переднавчання містить засвоєння мовних закономірностей і структур, перш ніж модель буде налаштована (фінтюнінг) для специфічних завдань.

Налаштування (фінтюнінг) для спеціалізованих завдань. Після попереднього тренування, XLM-R може бути налаштована на конкретні завдання, такі як класифі-

кація емоційного забарвлення (сентимент-аналіз), відповіді на питання, машинний переклад тощо.

Компонент "Класифікатор" використовується для валідації останньої версії класифікатора і застосування сентимент-аналізу на даних, згодом цей результат буде подано в інтерфейсі користувача, а також для оцінювання точності моделі на тестових даних. Існують різні метрики для оцінювання того, наскільки добре працюють моделі класифікації. Точність (accurasy), яка визначає кількість правильних передбачень у кожному прогнозі, є базовою метрикою, яка використовується для оцінювання моделі

$$Accurasy = P/N,$$

де: P – кількість текстів, за якими класифікатор прийняв правильне рішення; N – розмір навчальної вибірки.

Однак, цей показник значною мірою залежить від збалансованості даних. Тому, якщо дані не були збалансованими, можна використовувати показник F-міри, який є середнім гармонійним показником як точності (precision), так і повноти (recall), відповідно [24].

Precision використовується, коли небажаним є хибно-позитивний клас, recall, коли хибно-негативний. Precision і Recall розраховуємо за формулами:

$$Precision = TP/(TP + FP),$$
$$Recall = TP/(TP + FN),$$

де: TP – істинно-позитивне рішення; FP – хибно-позитивне рішення; FN – хибно-негативне рішення.

F-міра – загальна міра точності моделі, містить інформацію про точність і повноту вибраного алгоритму. Розраховуємо за формулою

$$F = (Precision \cdot Recall)/(Precision + Recall).$$

Отже, F-міра, Recall та Precision є головними метриками для оцінювання моделі.

Компонент "Інтерфейс користувача" для NLP-системи для сентимент-аналізу коментарів у соціальних мережах має бути добре розробленим та містити різні сервіси для забезпечення зручної та ефективної роботи з NLP-системою. Ось кілька важливих функцій, які можна використовувати:

1. Інтерфейс вибору компанії/продукту – дає змогу користувачам обирати компанію або її продукт.
2. Кнопка аналізу – додайте кнопку або елемент керування, який розпочне процес аналізу коментарів після введення тексту.
3. Графічні елементи для відображення результатів: відображення сентименту коментарів у вигляді графічних значків, графіків або кольорових показників (наприклад, позитивний – зелений, негативний – червоний, нейтральний – сірий).
4. Текстові результати: показувати сентимент коментарів у текстовому вигляді (наприклад, "Позитивний", "Негативний", "Нейтральний") разом із числовими показниками (наприклад, відсотком позитивних та негативних слів).
5. Графік часового ряду: якщо аналіз здійснюється для коментарів від користувачів упродовж певного періоду, можна використовувати графік часового ряду для відображення зміни сентименту з часом.
6. Сортування та фільтрація: додайте можливість сортування та фільтрації коментарів за сентиментом, датою, кількістю лайків й іншими параметрами.
7. Графік сутностей: якщо аналіз містить визначення ключових сутностей у тексті (наприклад, осіб або продуктів), то відобразіть їх на графіку або списку.
8. Інформаційні панелі: внесіть інформаційні панелі або віджети, які надають користувачам контекст про аналізовані дані, які алгоритми та методи використовуються для аналізу сентименту.

У нашому дослідженні було приділено особливу увагу верифікації і валідації отриманих результатів візуалізацією результатів сентимент-аналізу.

На кругових діаграмах показано результати валідації NLP-застосунку сентимент-аналізу відгуків користувачів послуг операторів мобільного зв'язку в Україні (рис. 8-11).

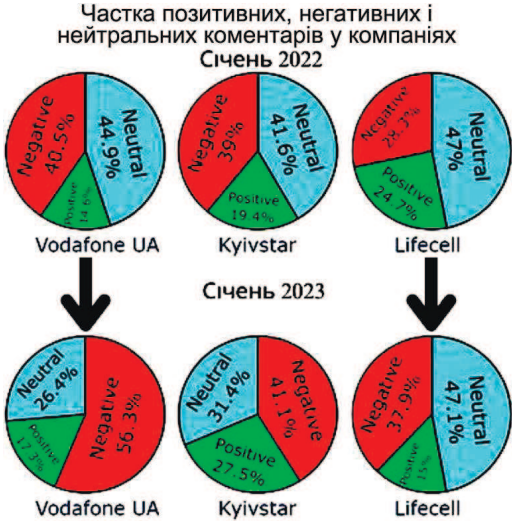

Рис. 8. Візуалізація результату сентимент-аналізу / Visualization of the result of sentiment analysis

На рис. 8 порівнюємо період до та після початку повномасштабного вторгнення. У відсотковому співвідношенні, у кожного оператора мобільного зв'язку стався приріст негативних коментарів. Далі розглянемо піки негативних коментарів і причини їх виникнення для кожного оператора (див. рис. 9-11).

Бачимо, що зміни в кількості негативних коментарів в компанії "Київстар" не змінились після початку вторгнення. Але в січні 2023 р. був найбільший пік 24 січня, порівняно з загальним трендом. Головною причиною є масові скарги користувачів через відсутність покриття і зв'язку через відключення електропостачання. Також головним фактором слугували масові повітряні тривоги і прильоти в регіонах, які є близькими до ураження. Саме з цих регіонів надійшла більшість скарг про тривалу відсутність зв'язку.

У ситуації з компанією "Лайфсел" бачимо загальний ріст кількості коментарів користувачів з року в рік і схожий загальний тренд негативних коментарів. Розглянемо детальніше негативні пікові значення в січні 2023 р. Бачимо 13, 17 і 25 січня. Здебільшого аналогічна ситуація, як і в "Київстар". Споживачі масово скаржаться на відсутність зв'язку в регіонах в період відсутності електроенергії по всій Україні. Також інша половина абонентів у ці дні скаржаться на стандартні проблеми, такі як малу швидкість мобільного Інтернету, відсутність покриття і періодичне зникнення мережі.

У ситуації з "Водафон Україна" бачимо загальний ріст коментарів в усіх трьох класах. У січні 2022 р. також були піки скарг, які частіше за все пов'язані з послугами 4G-мережі, покриттям мережі та тарифами і складністю роботи з додатком. Проте у 2023 р. бачимо, що кількість негативних коментарів переважає кількість інших класів в усіх днях. Маємо такі пікові дати, це 20, 25 і 28 січня 2023 року. 20 січня найбільше негативних коментарів за цей період з'явилися після новини

про кількість нових базових станцій від оператора, внаслідок чого люди стали писати більше про відсутність покриття, Інтернету і зв'язку в період від'єднань. Як і в інших операторів, 24-25 січня має пікові значення після обстрілів. На 28 січня переважна кількість негативних коментарів припадає на зміну тарифних планів з переважним збільшенням абонплати. Також інша частина скаржиться на відсутність мережі.

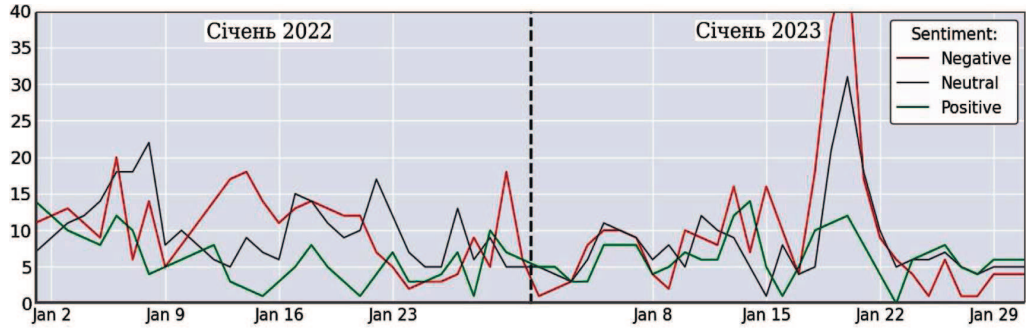

Рис. 9. Динаміка сентимент-аналізу для компанії "Київстар" / Dynamics of sentiment analysis for Kyivstar

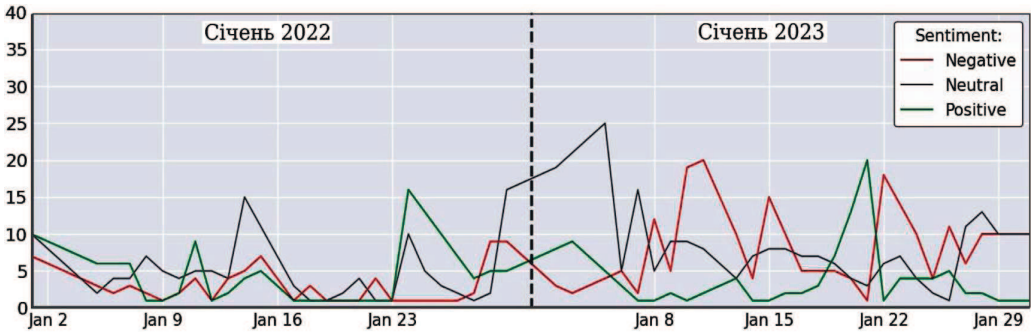

Рис. 10. Динаміка сентимент-аналізу для компанії "Лайфсел" / Dynamics of sentiment analysis for Lifecell

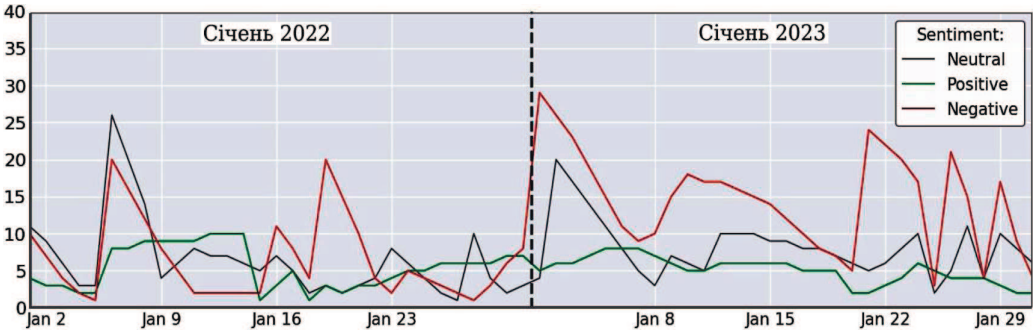

Рис. 11. Динаміка сентимент-аналізу для компанії "Водафон Україна" / Dynamics of sentiment analysis for Vodafone Ukraine

Отже, бачимо, що період від'єднань електроенергії вплинув на всіх операторів мобільного зв'язку без винятку. В усіх трьох операторів було помічено в період 24-25 січня (після обстрілів регіонів України) сплеск негативних коментарів на проблеми з покриттям і роботою мережі в період аварійних від'єднань. В інші пікові дні головними проблемами залишаються покриття, стабільність мережі в регіонах і швидкість мобільного зв'язку. Аналогічні проблеми було зазначено в коментарях 2022 р. до початку повномасштабного вторгнення, але кількість відгуків в усіх операторів збільшилась, як і кількість негативних коментарів.

Отже, концептуальна модель NLP-системи та NLP-застосунок сентимент-аналізу для вивчення відгуків користувачів послуг операторів мобільного зв'язку в Україні базується на принципах системної інженерії та бізнес-профілі Еріксона-Пенкера, що забезпечує структурований підхід до розроблення такої NLP-системи. Використання концептуальної моделі NLP-системи дає змогу формалізувати класи сутностей NLP-системи заданого призначення: проблеми, завдання та задачі, мету розроблення та застосування, необхідні ресурси і бізнес-правила функціонування, процеси функціонування та відношення між цими класами сутностей.

Одним із ключових елементів дослідження є застосування багатомовної моделі XLM-R для оброблення текстових даних різними мовами, враховуючи українську. Порівняно з іншими підходами, такими як окремі моделі для кожної мови чи використання машинного перекладу, багатомовна модель продемонструвала кращу продуктивність та точність під час класифікації сентименту відгуків користувачів.

Окрім цього, у дослідженні було приділено належну увагу питанням верифікації та валідації системи. Процес верифікації забезпечив відповідність результатів розроблення встановленим вимогам, тоді як валідація підтвердила можливість застосування системи за реальних умов експлуатації.

Одним із можливих напрямів подальших етапів дослідження може бути розширення моделі та застосування NLP-системи для аналізу відгуків користувачів у різних галузях або регіонах. Це дасть змогу перевірити масштабованість та гнучкість запропонованого рішення, а також виявити потенційні проблеми чи обмеження, які можуть виникнути в разі застосування NLP-системи в нових контекстах.

Важливо продовжувати вдосконалення методів та алгоритмів сентимент-аналізу, особливо в контексті му-

льтимодальних даних, таких як відео та зображення. З розвитком соціальних медіа та популярності візуального контенту, здатність аналізувати сентимент з різних джерел даних стає дедалі актуальнішою.

Результати цього дослідження демонструють можливість розроблення валідних NLP-застосунків з низькою повною вартістю володіння. Актуальність нашої розробки полягає у тому, що концептуальна модель NLP-системи може бути застосована для розроблення будь-яких NLP-застосунків сентимент-аналізу, а побудований на її основі NLP-застосунок сентимент-аналізу відгуків користувачів послуг операторів мобільного зв'язку в Україні є прикладом спеціалізованого рішення для аналізу відгуків користувачів послуг мобільних операторів України, що зосереджується на розробленні та імплементації штучного інтелекту та NLP-модулів, які враховують специфіку української мови і регіональних особливостей споживачів цих послуг. Важливою особливістю запропонованого NLP-застосунку може стати спрямованість на глибший аналіз контексту відгуків україномовних текстів, що не завжди можливо реалізувати засобами потужних універсальних інструментів Brand24 та MonkeyLearn.

Обговорення результатів дослідження. Тема роботи спрямована на проведення пошукових етапів дослідження з розроблення концептуальної моделі NLP-систем та її практичного застосування для розроблення методів системної інженерії NLP-систем та NLP-застосунків сентимент-аналізу. Це потребує аналізу та систематизації всіх основних сутностей оброблення природної мови під час сентимент-аналізу. Під час сентимент-аналізу (визначення настрою тексту) комп'ютер шукає п'ять головних сутностей: об'єкт оцінки, його конкретну властивість (аспект), саму оцінку (тональність), автора думки та час, коли її висловили. Ці елементи допомагають програмам зрозуміти, що саме подобається або не подобається користувачам у відгуках чи коментарях.

У фундаментальній роботі [25] автори досліджують та аналізують наявні підходи до оброблення природної мови (NLP), методи і моделі комп'ютерної лінгвістики та розпізнавання мовлення. У 3-му виданні (драфт 2023-2025 рр.) основну увагу приділяють нейромережевим методам, великим мовним моделям та сучасним алгоритмам NLP. Зокрема, це такі категорії, як: слова та токени, відстань редагування, N-грамні мовні моделі, логістична регресія та класифікація тексту, вбудовування, нейронні мережі, великі мовні моделі, трансформери, після навчання: налаштування інструкцій, вирівнювання та обчислення під час тестування, моделі маскованої мови, пошук інформації та генерування з доповненим пошуком, машинний переклад, RNN та LSTM, фонетика та вилучення ознак мовлення, автоматичне розпізнавання мовлення, перетворення тексту на мовлення, послідовне маркування частин мови та іменованих сутностей, контекстно-вільні граматики та розбір складових частин, розбір залежностей, вилучення інформації: відношення, події та час, маркування семантичних ролей та структура аргументів, лексикони для настроїв, афектів та конотацій, розв'язання кореферентності кореференцій та зв'язування сутностей, зв'язність дискурсу, розмова та її структура та інші сутності. Практичне застосування цієї множини NLP-інструментів потребує застосування системної інженерії та концептуального моделювання NLP-систем та NLP-застосунків.

У роботі [5] автори досліджують проблему застосування семантичних мереж. Автори спираються на поняття семантичної мережі – як формалізоване подання структурованих знань у вигляді графа. Такі мережі складаються з вузлів, які представляють поняття (наприклад, слова чи фрази), та посилань, які представляють семантичні відношення між вузлами. Посилання – це кортежі формату (джерело, семантичне відношення, місце призначення), які кодують знання. Автори наголошують, що попередні дослідження семантичних мереж зосереджувалися на кількох основних властивостях та спиралися на численні набори даних зі змішаними семантичними зв'язками, які детально аналізуються в розділі "Суміжні роботи" цієї статті. Автори прийшли до висновку, що не було проведено систематичного та комплексного аналізу топологічних властивостей семантичних мереж на рівні семантичних зв'язків.

Отже, залишається актуальною і проблема реалізації семантичних мереж для розроблення NLP-систем на множині даних та знань конкретної предметної області. У роботі [45] автори аналізують актуальну проблему машинного перекладу на основі великих мовних молей (LLM), зокрема, чи може застосування (LLM) та мультимодальні підходи у машинному перекладі (MT) покращити адекватність і точність перекладу. Результати подібних досліджень опубліковано й у роботі [30], де автори наводять структурне та динамічне подання системи машинного нейронного перекладу та її практичну реалізацію.

У фундаментальній роботі [43] автори зібрали, розглянули та дослідили сутності та інструменти концептуального моделювання за останні 50 років, що були опубліковані у понад 5300 статях, зібраних з 35 міждисциплінарних журналів та конференцій. Обговорюють важливу роль, яку концептуальне моделювання має відігравати у цифровому світі, що розвивається, та пропонують майбутні напрями досліджень. Автори встановили три основні напрями розвитку концептуального моделювання – дослідження та розроблення моделей концептуальних процесів, моделювання зв'язків між сутностями концептуальних моделей та онтологія концептуального моделювання. Автори вказують, що істотно менше досліджень проводиться в таких важливих галузях концептуального моделювання, як соціальне моделювання намірів, цілей та цінностей. Проте, потреба в додаткових дослідженнях у цих напрямах зростає, що підтверджується такими випадками застосування, як моделювання етичної поведінки при застосуванні штучного інтелекту або моделювання складних адаптивних систем. Автори вказують на те, що концептуальне моделювання стає життєво важливим інструментом досліджень і розробок інформаційно-комунікаційних систем та застосунків.

У роботі [38], як і у багатьох інших публікаціях, дослідники вивчають актуальну проблему застосування методів та засобів оброблення природної мови (NLP) для аналізу та генерування навчальних матеріалів українською мовою. Автори встановили, що наразі у науковому обігу є тільки обмежені ресурси української мови, відсутні потужні галузеві набори даних для досліджень та для тренування моделей штучного інтелекту. Автори вказують на те, що застосування таких ресурсів призводить до низької якості результатів, отриманих від NLP-моделей. Автори порівнюють результати застосування

мовних моделей OpenAI та BERT, зокрема їх точність, контекстуальність та адаптивність до української мови. Автори наголошують на необхідності розроблення спеціалізованих NLP-моделей, адаптованих до особливостей української мови і продукування якісних навчальних даних.

У роботі [49] автори запропонували нову архітектуру нейронної мережі – Transformer, де замість рекурентних зв'язків (RNN) та згорткових шарів (CNN) застосовують механізм уваги (Self-Attention). Це дало змогу перейти від послідовного оброблення слів до паралельного, що істотно пришвидшує навчання та моделювання. Істотною перевагою застосування трансформерів для оброблення природної мови є зменшення потреб в обчислювальних ресурсах. Архітектура Transformer стала основою для розроблення великих мовних моделей, зокрема BERT, GPT та Claude.

Отже, внаслідок обговорення результатів дослідження було підтверджено доцільність проведення пошукових етапів дослідження з розроблення концептуальної моделі NLP-систем та її практичного застосування для розроблення методів системної інженерії NLP-систем та NLP-застосунків сентимент-аналізу, оскільки подані приклади результатів публікацій не повною мірою висвітлюють цей напрям етапів дослідження і практичних еапів розробок.

Концептуальне моделювання систем та застосунків з оброблення природної мови (NLP) – це процес створення абстрактного, структурованого подання мовних знань, семантичних зв'язків, правил інтерпретації людської мови, який слугує основою для проєктування моделей NLP-систем і застосунків. Концептуальне моделювання описує структурне і динамічне подання систем та застосунків з оброблення природної мови і дає змогу інтегрувати наявні та нові наукові та інженерні рішення ще на етапі проєктування систем та застосунків з оброблення природної мови.

Отже, внаслідок виконаної роботи можна сформулювати такі наукову новизну та практичну значущість результатів дослідження.

Наукова новизна отриманих результатів дослідження – отримала подальший розвиток концептуальна модель NLP-систем, яка, на відміну від наявних, побудована на основі бізнес-профілю Еріксона-Пенкера, що дає можливість врахувати і подати всі необхідні і достані класи сутностей NLP-систем та відношення між ними на основі прийнятих аксіоматичних положень і понять класів "Проблема, завдання, задача інженерії NLP-систем", "Мета розроблення", "Процес функціонування", "Ресурси" та "Бізнес-правила функціонування".

Практична значущість результатів дослідження – концептуальну модель NLP-систем можна використати для системної інженерії NLP-застосунків будь-якого призначення та області застосування на множині заданих ресурсів та встановлених бізнес-правил у межах заданого буджету.

Висновки / Conclusions

Розроблено концептуальну модель NLP-системи для формалізації та подання NLP-застосунків з оброблення природної мови, що дало змогу отримати вербальну модель необхідної множини сутностей і процесів їх взаємодії задля виявлення недоліків і уточнення предмету дослідження. За результатами проведеного дослідження можна зробити такі основні висновки.

1. Запропонована концептуальна модель NLP-системи на основі бізнес-профілю Еріксона-Пенкера та системного підходу дає змогу врахувати і формалізувати всі необхідні класи сутностей NLP-систем та сформувати відношення між ними.
2. Запропонована модель NLP-застосунку сентимент-аналізу для вивчення відгуків користувачів мобільного зв'язку в Україні побудована на основі концептуальної моделі NLP-системи та багатомовної моделі XLM-R для оброблення текстових даних різними мовами, враховуючи українську. Порівняно з іншими моделями, такими як окремі моделі для кожної мови, чи використання машинного перекладу, багатомовна модель продемонструвала кращу продуктивність та точність під час класифікації сентименту відгуків користувачів.
3. Проведена валідація концептуальної моделі NLP-системи та NLP-застосунку сентимент-аналізу для вивчення відгуків користувачів операторів мобільного зв'язку в Україні, підтверджена результатами моделювання для трьох основних операторів мобільного зв'язку "Київстар", "Лайфсел" та "Водафон Україна".
4. Результати цього дослідження демонструють можливість розроблення валідних NLP-застосунків з низькою повною вартістю володіння.
5. Перспективи подальших етапів дослідження спрямовані на розроблення методик системної інженерії NLP-систем сентимент-аналізу відгуків користувачів продуктів, товарів та послуг галузевих виробників та постачальників в Україні.

References

1. Amazon Web Services, Inc. (n.d.). *Amazon Simple Storage Service Documentation*. Amazon. Retrieved May 11, 2026. URL: https://docs.aws.amazon.com/s3/
2. Anand, A. (2020, September 2). *Cross lingual models* (*XLM-R*). Medium. URL: https://medium.com/@aman.anand54321/cross-lingual-models-xlm-r-7d557302698b
3. Borthakur, D. (2026). *HDFS Architecture Guide*. Apache Hadoop. URL: https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html
4. Brand24. (n.d.). *Brand24 – #1AI social listening tool*. Brand24. URL: https://brand24.com/
5. Budel, G., Jin, Y., Van Mieghem, P., & Kitsak, M. (2023). Topological properties and organizing principles of semantic networks. *Scientific Reports*, *13*, article ID 11728. https://doi.org/10.1038/s41598-023-37294-8
6. Cer, D., Yang, Y., Kong, S.-y., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil, R. (2018). Universal Sentence Encoder. *arXiv preprint arXiv:1803.11175*. URL: https://arxiv.org/abs/1803.11175
7. Chaudhuri, K. D. (2022, March). *Building Naive Bayes Classifier from Scratch to Perform Sentiment Analysis*. Analytics Vidhya. URL: https://www.analyticsvidhya.com/blog/2022/03/building-naive-bayes-classifier-from-scratch-to-perform-sentiment-analysis/
8. Conneau, A., Khandelwal, Kartikay., & Goyal, Naman., et al. (2020). Unsupervised Cross-lingual Representation Learning at Scale. *arXiv.org*. https://doi.org/10.48550/arXiv.1911.02116
9. Curtis, J. (2025). *What is a Data Lake? – Introduction to Data Lakes and Analytics – AWS*. Amazon Web Services, Inc. URL: https://aws.amazon.com/what-is/data-lake/
10. Edmonds, Ph., & Agirre, E. (2008). Word sense disambiguation. Scholarpedia, *3*(7), article ID 4358. https://doi.org/10.4249/scholarpedia.4358
11. Eriksson, H.-Er. & Penker, M. (1999). Business Modeling with UML: *Business Patterns at work*, Wiley & Sons, Fall. URL: https://www.utm.mx/~caff/poo2/Business%20Modeling%20with%20UML.pdf
12. Expert.ai. (2022, April 26). *What semantic analysis means to natural language processing*. URL: https://www.expert.ai/blog/natural-language-process-semantic-analysis-definition/

13. Foster, P., & Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking. *OReilly Media*, Inc. URL: https://www.researchgate.net/publication/256438799_Data_Science_for_Business
14. Gandhi, R. (2018, June 7). *Support Vector Machine – Introduction to Machine Learning Algorithms*. Towards Data Science. URL: https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47
15. Gillham, J. (2025, August 20). *What are dialogue systems – How do they relate to AI content creation?* Originality.AI. URL: https://originality.ai/blog/what-are-dialogue-systems
16. Grytsiuk, P. Y., Ivanyshyn, A. V., & Hrytsiuk, Y. I. (2023). Quality assurance of software products in accordance with IEEE 730-2014 standard within the project implementation lifecycle. *Scientific Bulletin of UNFU*, 33(2), 101–117. https://doi.org/10.36930/40330214
17. Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., & Mathur, I. (2016). *Natural Language Processing: Python and NLTK*. Packt Publishing. URL: https://books.google.com.ua/books?id=0J_cDgAAQBAJ&pg=PA21&hl=uk&source=gbs_toc_r&cad=2#v=onepage&q&f=false
18. Harris, J. (2016, November 21). *The growing importance of big data quality*. SAS Blogs. URL: https://blogs.sas.com/content/datamanagement/2016/11/21/growing-import-big-data-quality/
19. Hashemi-Pour, C., & Barney, N. (2024, October 28). *What is named entity recognition* (*NER*)? [Definition]. TechTarget. URL: https://www.techtarget.com/whatis/definition/named-entity-recognition-NER
20. Hattenhauer, R. (2024). *ChatGPT & Co.: A Workbook for Writing, Research, Creating Images, Programming, and More* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781003503675
21. Hossain, B. A., Md. Mukta, S. H., Md. Islam, A., Zaman, A., & Schwitter, R. (2023). Natural Language – Based Conceptual Modelling Frameworks: State of the Art and Future Opportunities. *ACM Computing Surveys*, 56, 1, article ID 12, 26 p. https://doi.org/10.1145/3596597
22. Huang, H., & Zhang, B. (2009). *Text Segmentation. In: LIU, L., ÖZSU, M.T.* (*Eds*). *Encyclopedia of Database Systems*. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_421
23. Hugging, Face. (2026). *Hugging Face – The AI community building the future* [HomePage]. URL: https://huggingface.co/
24. Huilgol, P. (2024). Precision and Recall | Essential Metrics for Machine Learning (2023 Update). *Analytics Vidhya*. URL: https://www.analyticsvidhya.com/articles/precision-and-recall-in-machine-learning/
25. Jurafsky, D., & Martin, J. H. (2026). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition. Online manuscript released January 6, 2026. Prentice Hall. URL: https://web.stanford.edu/~jurafsky/slp3
26. Just, J. (2024). Natural language processing for innovation search – Reviewing an emerging non-human innovation intermediary. *Technovation, 129*, article ID 102883. https://doi.org/10.1016/j.technovation.2023.102883
27. Khani, F., & Ribeiro, M. T. (2023). *Collaborative Development of NLP models*. Hugging Face. URL: https://arXiv.org/pdf/2305.12219
28. Kirvan, P., Lutkevich, B., & Kiwak, K. (2024, November 20). *What is speech recognition?* [Definition from TechTarget]. Customer Experience. URL: https://www.techtarget.com/searchcustomerexperience/definition/speech-recognition
29. Lebedenko, N., Datsyshyn, K., Haladzhun, Z., Kunanets, N., Veretennikova, N., et al. (2024). Words with the Prefix De- in the Ukrainian Online Media: Structure, Semantics, Tonality. *Communication and Linguistics Studies*, *10*(2), 39–52. https://doi.org/10.11648/j.cls.20241002.13
30. Maslianko, P. P., & Sielskyi, E. P. (2021). Method of system engineering of neural machine translation systems. *KPI Science News*, № 2, 46–55. https://doi.org/10.20535/kpisn.2021.2.236939
31. Maslianko, P. R., & Maystrenko, O. S. (2008). The system engineering of organizational system informatization projects. *KPI Science News*, *6*, 4–42. URL: https://ela.kpi.ua/handle/123456789/36036
32. Maslianko, P., & Sielskyi, Y. (2021). Data Science – Definition and Structural Representation. *System Research & Information Technologies*, № 1, article ID 61–78. https://doi.org/10.20535/SRIT.2308-8893.2021.1.05
33. Meta Platforms, Inc. (n.d.). *Meta developer documentation | Meta APIs, SDKs & guides*. Meta for Developers. Retrieved May 11, 2026. URL: https://developers.facebook.com/docs/
34. Microsoft. (2026, April 15). *Use the Azure Data Lake Storage URI* [Documentation]. Microsoft Learn. URL: https://learn.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction
35. MIT Technology Review Insights. (2026, April 14). *Redefining the future of software engineering: How agentic AI will change the way software is developed and managed*. [Report]. URL: https://www.technologyreview.com/2026/04/14/1134397/redefining-the-future-of-software-engineering/
36. MonkeyLearn. (2026). *Text analysis platform to unlock insights from customer feedback*. monkeylearn.com. URL: https://www.medallia.com/platform/text-analytics/
37. Ni, J., Young, T., Pandelea, V., et al. (2023). Recent advances in deep learning based dialogue systems: a systematic survey. *Artif Intell Rev,* 56, 3055–3155. https://doi.org/10.1007/s10462-022-10248-8
38. Patsai, B., Nechyporuk, I., & Kovtun, A. (2025). Natural language processing in Ukrainian: challenges and prospects for the use of artificial intelligence in education. *Digital Economy and Economic Security, 1*(16), 172–179. https://doi.org/10.32782/dees.16-26
39. Rosu, R., Stefan Stoica, A., Popescu, P. St., & Mihăescu, M. Cr. (2020). NLP based Deep Learning Approach for Plagiarism Detection. *International Journal of User-System Interaction*, *13*(1), article ID 48–60. https://doi.org/10.37789/ijusi.2020.13.1.4
40. Savchenko, Ye. A., & Savchenko, M. Yu. (2020). The Transfer Learning Task as the Means of Metalearning Tasks Solution. *Control Systems and Computers*, *2*, 41–54. https://doi.org/10.15407/csc.2020.02.041
41. Shikha. (2025, May 26). *Data lake vs. data warehouse: Whats the difference?*. Analytics Vidhya. URL: https://www.analyticsvidhya.com/blog/2023/02/a-comprehensive-guide-to-data-lake-vs-data-warehouse/
42. Siva, G. (2021, November 2). *BERT – Bidirectional Encoder Representations from Transformer*. Medium. URL: https://gayathri-siva.medium.com/bert-bidirectional-encoder-representations-from-transformer-8c84bd4c9021
43. Storey, V. C., Lukyanenko, R., & Castellanos, A. (2023). Conceptual modeling: Topics, themes, and technology trends. *ACM Computing Surveys*, *55*(14s), article ID 317, 1–38. https://doi.org/10.1145/3589338
44. Stryker, C., & Holdsworth, J. (2026). *What is NLP* (*Natural Language Processing*)? IBM. URL: https://www.ibm.com/think/topics/natural-language-processing
45. Suram, P., Banik, D., & Sharma, A. K. (2026). Large language model based machine translation for universal multilingual understanding and translation quality enhancement. *Discover Computing*, *29*, article ID 217. https://doi.org/10.1007/s10791-026-10027-x
46. Thanaki, J. (2017). Python Natural Language Processing: Advanced machine learning and deep learning techniques for natural language processing. Packt Publishing. URL: https://dl.acm.org/doi/book/10.5555/3164989#cited-by-sec
47. The Stanford Natural Language Processing Group. (2025). *Machine translation*. URL: https://nlp.stanford.edu/projects/mt.shtml
48. Torskyi, O. I., & Hrytsiuk, Y. I. (2025). Application of machine learning to enhance the efficiency of automated software testing. *Scientific Bulletin of UNFU*, 35(4), 142–149. https://doi.org/10.36930/40350416
49. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems, 30*, 5998–6008. Curran Associates, Inc. https://doi.org/10.48550/arXiv.1706.03762
50. Zhang, L., & Sun, J. T. (2009). Text generation. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of database systems* (pp. 3065–3070). Springer. https://doi.org/10.1007/978-0-387-39940-9_416

P. P. Maslianko, Yu. S. Mysak
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

CONCEPTUAL MODEL OF NLP SYSTEM AND NLP APPLICATION SENTIMENT ANALYSIS OF USER FEEDBACK OF MOBILE OPERATORS SERVICES IN UKRAINE

Natural Language Processing (NLP) tasks and objectives are intended to tackle a broad spectrum of scientific and practical problems. The paper identifies that organising these tasks outlines specific areas for research, theoretical, and applied development in NLP. These include widely discussed topics among engineers and R&D companies, such as text segmentation, morphological analysis, syntactic and semantic analysis, machine translation, text generation, sentiment analysis, selective search, information analysis, and a multitude of additional applications. The research and development of axiomatic tools for representing NLP systems, based on a systems approach and systems engineering, is considered a relevant issue. A scientifically validated conceptual model for an NLP system is the outcome of this approach. We employed the method of Systems Engineering of Information and Communication Systems to develop this model, specifically the Ericsson-Penker business profile, which facilitates the formalisation of the NLP system by defining a set of specific, named entity classes. It also enables the identification of the systems challenges, objectives, and tasks, as well as its purpose, resources, processes, and the business rules governing its operation. The proposed conceptual model of the NLP system offers a systematic arrangement of the essential and adequate set of entity classes along with the relationships that exist between them, facilitating the creation of both a structural and dynamic representation of the systems architecture, while also enabling the identification of the internal organisation of components and the interaction interfaces necessary for executing specific application tasks. The paper illustrates a practical example of conceptual modelling applied to the development of an NLP system aimed at analysing the tone of responses from users of mobile operators services in Ukraine, and it also verifies and validates the operation of the developed application. Future research prospects involve enhancing the conceptual model and formulating methodologies for the systems engineering of NLP systems to address particular categories of natural language processing and analysis challenges.

Keywords: Natural Language Processing; Erickson-Penker business profile; systems engineering of NLP systems; representation of NLP systems; conceptual model of NLP system.