

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ

УДК 004

ТЕЗИ ДОПОВІДЕЙ

МІЖНАРОДНОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ
“СУЧАСНІ ІНФОРМАЦІЙНІ СИСТЕМИ ТА ТЕХНОЛОГІЇ
В ЦИФРОВОМУ СУСПІЛЬСТВІ”

10 - 11 КВІТНЯ 2025 Р.

ABSTRACTS OF REPORTS

INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE
"MODERN INFORMATION SYSTEMS AND TECHNOLOGIES
IN THE DIGITAL SOCIETY"

APRIL 10 - 11, 2025

Харків
2025

Jean Monnet Activities



ТЕХНОЛОГІЇ ОБРОБКИ ПРИРОДНОЇ МОВИ: АВТОМАТИЗАЦІЯ АНАЛІЗУ ТЕКСТОВИХ ДАНИХ

Сьогодні у світі щодня генеруються петабайти текстової інформації, а обсяг неструктурованих даних зростає зі швидкістю близько 55-65% щороку. Це створює потребу в розробці ефективних технологій автоматизованої обробки та аналізу текстових даних. Технології обробки природної мови (NLP) представляють собою міждисциплінарну галузь, що поєднує лінгвістику, комп'ютерні науки та штучний інтелект для вирішення завдань розуміння, інтерпретації та генерації людської мови комп'ютерними системами. Для досягнення цієї мети в NLP використовується широкий спектр методів і технік, починаючи з базової обробки тексту і закінчуючи складними моделями машинного навчання.

Метою даного дослідження є аналіз сучасних підходів до автоматизації обробки текстових даних, оцінка їх ефективності та визначення перспективних напрямків розвитку.

Розвиток технологій обробки природної мови пройшов декілька ключових етапів: правилі системи, статистичні методи, епоха векторних представлень, глибоке навчання [1], архітектури з механізмами уваги започаткували нову еру в NLP. Трансформери та їх реалізації, такі як BERT [2], GPT, RoBERTa, T5, стали домінуючою архітектурою, що забезпечує контекстне розуміння тексту, демонструють неперевершені можливості в генерації мови та веденні багаторандомових розмов. Великі мовні моделі масштабували архітектури трансформерів до сотень мільярдів параметрів, що дозволило системам демонструвати здатність до розуміння контексту, міркування та генерування когерентного тексту (GPT-3, GPT-4, LLaMA [3], Claude).

Попередня обробка тексту є критичним етапом у побудові ефективних NLP-систем і зазвичай включає: нормалізацію тексту, токенизацію, вилучення стоп-слів, лематизацію/сте́мінг, part-of-speech tagging, векторизацію.

Сучасні NLP-технології дозволяють ефективно вирішувати широкий спектр завдань таких як: класифікація текстів, аналіз тональності (sentiment analysis), вилучення іменованих сутностей (NER), автоматичне реферування, машинний переклад, генерація тексту, відповіді на запитання, виявлення спаму, чат-боти та віртуальні асистенти.

У рамках нашого дослідження було проведено порівняльний аналіз різних архітектур для задачі класифікації українськомовних текстів. Експерименти проводилися на корпусі українських новин. Моделі на базі трансформерів демонструють

найвищі показники якості, особливо коли вони спеціально адаптовані для української мови.

Незважаючи на значний прогрес, технології NLP стикаються з низкою викликів. Одним з головних є складність та неоднозначність людської мови. Мовна дисперсія – більшість моделей оптимізовані для англійської мови. Обчислювальна складність – сучасні моделі вимагають значних обчислювальних ресурсів для навчання та використання. Інтерпретованість – нейронні моделі часто функціонують як "чорні скриньки", що ускладнює розуміння їхніх рішень та виправлення помилок. Помилки та упередження – моделі можуть наслідувати та підсилювати упередження, присутні в навчальних даних. Конфіденційність даних – використання реальних текстових даних для навчання моделей створює ризики для приватності.

Розвиток NLP-технологій піднімає ряд етичних питань таких як: конфіденційність та захист даних, прозорість та пояснюваність, упередження та дискримінація, дезінформація. Для мінімізації цих ризиків використовуються підходи відповідального ІІІ (Responsible AI).

Технології обробки природної мови переживають період швидкого розвитку, обумовлений прогресом у галузі архітектур нейронних мереж та доступністю обчислювальних ресурсів. Сучасні моделі на базі трансформерів та великі мовні моделі демонструють вражаючі результати в різноманітних завданнях аналізу тексту. Майбутнє NLP лежить у розвитку більш ефективних, інтерпретованих та етично відповідальних моделей, що зможуть якісно працювати з різними мовами та предметними областями.

Список літератури

1. Y. Goldberg, Neural Network Methods for Natural Language Processing. Synthesis Lectures on Human Language Technologies, 2017.
2. J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of NAACL-HLT 2019, 2019. [Online]. Available: <https://aclanthology.org/N19-1423/>
3. V. Veeramachaneni, Large Language Models: A Comprehensive Survey on Architectures, Applications, and Challenges. Advanced Innovations in Computer Programming Languages, 7(1), pp. 20–39. 2025. [Online]. Available: <https://doi.org/10.5281/zenodo.14161613>

СЕКЦІЯ 2. Системи та технології інтелектуальної обробки даних

Brynza Natalia OPTIMISATION OF WEB PAGE LOADING SPEED	50
Donsky Dmytro, Manakov Volodymyr METHOD OF RANKING ELECTRONIC TEXT DOCUMENTS BY RELEVANCE TO A QUERY	51
Hryshko Andrey, Udovenko Serhiy PREDICTING THE TREND OF REINFORCEMENT SIGNALS IN INTELLIGENT DIGITAL CONTROL SYSTEMS	52
Ivanisenko Ihor LOAD BALANCING SYSTEM IN MULTISERVICE COMPUTER NETWORKS.....	53
Teslenko Oleh STAGES OF RESEARCHING THE EFFECTIVENESS OF MODERN METHODS OF OPTIMISING WEB PAGE LOADING SPEED	54
Безкоровайний Володимир, Драз Оксана ВІДНОВЛЕННЯ МЕРЕЖІ ТЕРМІНАЛІВ ЛОГІСТИЧНИХ МЕРЕЖ ЗА НАСЛІДКАМИ НАДЗВИЧАЙНИХ СИТУАЦІЙ	55
Болотніков Захарій , Чала Лариса МУЛЬТИМОДАЛЬНА СИСТЕМА ОБРОБКИ ТА АНАЛІЗУ КОРОТКИХ НОВИННИХ ВІДЕО	56
Кобзев Ігор, Дроботенко Владислав ЗАСОБИ ЗАХИСТУ ВЕБ-САЙТІВ, СТВОРЕНИХ НА СИСТЕМІ УПРАВЛІННЯ КОНТЕНТОМ WORDPRESS.....	57
Кобзев Ігор, Горелов Юрій НАПРЯМКИ ВИКОРИСТАННЯ BIG DATA ПІД ЧАС ВІЙНИ В УКРАЇНІ	58
Купін Андрій, Голівер Владислав ПАРАЛЕЛЬНА ОБЧИСЛЮВАЛЬНА АРХІТЕКТУРА НА ОСНОВІ ГРАФІЧНИХ ПРОЦЕСОРІВ ТИПУ CUDA	59
Лебідь Вадим, Гриньова Олена ІНТЕЛЕКТУАЛЬНА СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ОПТИМІЗАЦІЇ РОЗМІЩЕННЯ РЕКЛАМИ В ЦИФРОВИХ ЗМІ.....	60
Передрій Олена, Магтогян Анастасія ТЕХНОЛОГІЇ ОБРОБКИ ПРИРОДНОЇ МОВИ: АВТОМАТИЗАЦІЯ АНАЛІЗУ ТЕКСТОВИХ ДАНИХ	61
Пиріг Наталія, Чала Лариса ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ВИЗНАЧЕННЯ ПРІОРИТЕТНОСТІ РЕКОНСТРУКЦІЇ БУДІВЕЛЬ В УМОВАХ ПОСТВОЄННОГО ВІДНОВЛЕННЯ	62
Суворов Роман, Чала Лариса ВИЗНАЧЕННЯ АВТЕНТИЧНОСТІ КОНТЕНТУ З ВИКОРИСТАННЯМ НЕЙРОННИХ МЕРЕЖ	63
Тютюник Ольга, Носань Юрій ОСОБЛИВОСТІ ФУНКЦІОНУВАННЯ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ В УМОВАХ НЕВИЗНАЧЕНОСТІ ІНФОРМАЦІЇ ПРО ПСИХОЛОГІЧНИЙ СТАН ЛЮДИНИ.....	64
Тютюник Вадим, Усачов Дмитро УДОСКОНАЛЕННЯ ПРОЦЕДУРИ ВИЯВЛЕННЯ ТА ІДЕНТИФІКАЦІЇ ЗА АМПЛІТУДНО-ЧАСТОТНИМИ ХАРАКТЕРИСТИКАМИ АКУСТИЧНИХ СИГНАЛІВ БЕЗПЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ	65
Угрімова Катерина, Кобзев Ігор ПІДХОДИ ДО РОЗРОБКИ МОБІЛЬНИХ ДОДАТКІВ	66

ТЕЗИ ДОПОВІДЕЙ
Міжнародної науково-практичної конференції
“Сучасні інформаційні системи та технології в
цифровому суспільстві”
10 - 11 квітня 2025 р.

Abstracts of reports
International scientific and practical conference
"Modern information systems and technologies
in the digital society"
April 10 - 11, 2025

Відповідальний за випуск: Д.О. Бондаренко

Комп'ютерна верстка: Д.Ю. Голубничий