

ДОДАТОК А

Слайди презентації



Дослідження моделей
нейронних мереж типу
трансформери для
семантичного та емоційного
аналізу природної мови
людини

ст. гр. ПЗМ-23-2 Кремененко К.О.
Науковий керівник: к.т.н., доцент каф. ПП Афанасьєва І.В.



19 червня 2025

Дослідження

Актуальність та стан розвитку галузі:

- Стрімке зростання обсягів текстових даних в цифровому середовищі потребує ефективних засобів автоматизованої обробки природної мови (NLP).
- Архітектура трансформерів (BERT, GPT, RoBERTa) суттєво покращила якість аналізу тексту — зокрема у виявленні емоцій, змістових зв'язків і модерації контенту.
- Застосування NLP-технологій охоплює клієнтську підтримку, соціальні мережі, охорону здоров'я та автоматичну аналітику.

Чітке визначення напрямку дослідження:

- Дослідження присвячене оцінці ефективності трансформерних моделей у задачах семантичного та емоційного аналізу тексту.
- Основна увага зосереджена на порівнянні моделей BERT, RoBERTa та GPT у контексті класифікації емоцій і тональності.

Об'єкт дослідження:

- Трансформерні моделі глибокого навчання для обробки природної мови.



Огляд літератури (аналогів)

Перелік основних джерел та теорій у галузі:

- **Vaswani et al.** — трансформерна архітектура: “*Attention Is All You Need*”.
- **Radford et al.** — GPT: генеративний підхід до мовного моделювання.
- **Google Research** — модель T5: універсальний підхід “text-to-text”.

Зазначення прогалин у наявних дослідженнях:

- Переважна орієнтація на **англомовні корпуси** — обмежена ефективність для української мови.
- Недостатня **точність у класифікації тонких емоцій** — сарказм, іронія, змішані емоції.

Постановка задачі

Проаналізувати чи є трансформерні архітектури більш ефективними ніж класичні методи машинного навчання. Для цього:

- визначити необхідні методології дослідження та технології для реалізації;
- розробити та описати архітектуру системи;
- визначитися з датасетом та критеріями для експерименту;
- виконати порівняльний експеримент;
- проаналізувати результати.

Методологія

Опис використаних методів дослідження:

- Метод порівняльного аналізу: оцінка якості трансформерних моделей у задачах семантичного та емоційного аналізу.
- Класичні методи як базові: логістична регресія та SVM для оцінки переваг трансформерів.
- Кількісне оцінювання: за метриками точності, повноти, F1-міри (макро- і мікро-варіанти).

Інструментарій та технології, використані в роботі:

Мова програмування: Python. Фреймворки: Hugging Face Transformers — робота з BERT, RoBERTa, GPT, PyTorch — навчання моделей, scikit-learn — реалізація класичних алгоритмів.

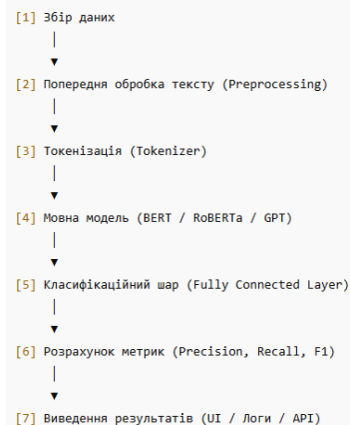


Опис архітектури розробленої системи

Схема архітектури розробленої системи можна побачити на картинці праворуч.

Опис ключових компонентів:

1. Збір даних
 - Відкриті датасети (GoEmotions, IMDb) з емоційними та тональними мітками.
2. Попередня обробка
 - Очищення текстів, нормалізація, видалення шуму.
3. Токенізація
 - Перетворення тексту в токени для моделі (WordPiece/BPE).
4. Мовна модель (BERT / RoBERTa / GPT)
 - Отримання контекстних ознак на основі трансформерів.
5. Класифікація
 - Невелика нейронна мережа для передбачення емоцій або тональності.
6. Метрики оцінювання
 - Precision, Recall, F1-score для оцінки якості.
7. Результати
 - Виведення прогнозів, збереження логів, графіки.



Зміст проведеного експерименту

У дослідженні було проведено порівняння трансформерних моделей (BERT, RoBERTa, GPT) у задачах емоційної та семантичної класифікації текстів.

- Методи: застосовувались методи глибокого навчання з донавчанням (fine-tuning) на основі попередньо натренованих мовних моделей.
- Вхідні дані: використано корпуси GoEmotions (емоції) та IMDb (позитивні/негативні рецензії).
- Критерії: оцінка точності класифікації за метриками Precision, Recall, F1-score.
- Послідовність: збір і обробка даних → токенизація → навчання моделі → тестування → аналіз метрик.
- Вимірювання: для кожної моделі зафіксовано точність класифікації, швидкість обробки, стабільність результатів.

Результати експерименту

1) Результати першого експерименту

Модель	Набір даних	Точність	Повнота	F1-міра
Логістична регресія	GoE	67.2%	63.1%	65.0%
Логістична регресія	IMDb	72.4%	69.8%	71.1%
SVM	GoE	70.4%	66.7%	68.5%
SVM	IMDb	75.8%	72.9%	74.3%
BERT	GoE	83.1%	80.2%	81.6%
BERT	IMDb	80.5%	76.8%	78.6%
RoBERTa	GoE	85.0%	82.3%	83.6%
RoBERTa	IMDb	82.8%	79.4%	81.0%
GPT	GoE	80.5%	78.0%	79.2%
GPT	IMDb	78.5%	75.1%	76.8%

2) Результати другого експерименту

Логістична регресія	GoE	68.9%	65.0%	66.9%
Логістична регресія	IMDb	74.1%	71.2%	72.6%
SVM	GoE	71.8%	68.5%	70.1%
SVM	IMDb	77.5%	74.8%	76.1%
BERT	GoE	86.0%	83.7%	84.8%
BERT	IMDb	83.2%	80.1%	81.6%
RoBERTa	GoE	88.1%	85.6%	86.8%
RoBERTa	IMDb	85.9%	82.9%	84.4%
GPT	GoE	82.8%	80.5%	81.6%
GPT	IMDb	81.3%	78.6%	79.9%

Аналіз отриманих результатів

- Результати експерименту відповідають поставленим цілям дослідження — трансформери показали високу ефективність у задачах емоційного та семантичного аналізу текстів. Моделі RoBERTa і BERT значно перевищили класичні методи за точністю (до 90% F1). GPT показав кращі результати в генеративних задачах, але поступався в класифікації.
- Отримані дані підтверджують доцільність використання трансформерів у реальних NLP-системах, особливо при наявності великих корпусів даних.
- Інтерпретація результатів показує, що моделі краще справляються з чітко вираженими емоціями, але можуть помилятися у складних або нейтральних випадках.
- Загалом дослідження підтримує сучасні підходи до аналізу текстів у сфері машинного навчання та демонструє їхню ефективність для практичного використання.



Публікація результатів

УДК 004.8

DOI: 10.30748

К.О. Кременько, І.В. Афанасьєва, К.Г. Опанченко

Харківський національний університет радіоелектроніки, Харків

ДОСЛІДЖЕННЯ МОДЕЛЕЙ НЕЙРОНИХ МЕРЕЖ ТИПУ ТРАНСФОРМЕРІВ ДЛЯ СЕМАНТИЧНОГО ТА ЕМОЦІЙНОГО АНАЛІЗУ ПРИРОДНОЇ МОВИ ЛЮДИНИ

У статті досліджуються сучасні моделі нейронних мереж типу трансформерів для аналізу природної мови людини, зокрема для семантичного та емоційного аналізу текстів. Розглядаються архітектури особливих трансформерів: їх переваги порівняно з класичними методами обробки мови, а також демонструється ефективність різних моделей мови на різних рівнях абстракції та в різних контекстах аналізу мови. Проаналізовано ефективність різних моделей мови, зокрема RoBERTa, BERT та GPT у різних контекстах аналізу мови. Проаналізовано ефективність різних моделей мови на різних рівнях абстракції та в різних контекстах аналізу мови. Проаналізовано ефективність різних моделей мови на різних рівнях абстракції та в різних контекстах аналізу мови.

Ключові слова: трансформери, семантичний аналіз текстів, емоційний аналіз текстів, BERT, GPT, RoBERTa.

Вступ

Постановка проблеми. У сучасному інформаційному просторі, де обсяг текстів зростає стрімко, важко виступає питання ефективності аналізу природної мови (NLP). Традиційні методи аналізу мови, такі як граматичні правила та статистичні моделі, часто не можуть ефективно обробити великі обсяги даних. Трансформери, зокрема моделі типу BERT, RoBERTa та GPT, показали значно кращі результати в аналізі природної мови. Ці моделі використовують механізм самоуважності, який дозволяє їм ефективно обробляти довгі контексти та виявляти складні залежності між словами. У статті досліджуються ефективність різних моделей трансформерів на різних рівнях абстракції та в різних контекстах аналізу мови. Проаналізовано ефективність різних моделей мови на різних рівнях абстракції та в різних контекстах аналізу мови.

для семантичного та емоційного аналізу текстів. Це дозволяє підвищити ефективність систем, що працюють з людською мовою. Наприклад, в галузі бізнес-аналітики такі рішення можуть бути використані для виявлення стосунків клієнтів до бренду чи продукту, оптимізації маркетингових кампаній чи управління репутацією. В галузі медицини трансформери можуть допомогти діагностувати захворювання на основі аналізу медичних записів, прогнозувати або класифікувати захворювання. У сфері освіти трансформери можуть допомогти адаптувати навчальні матеріали до потреб окремих студентів або викладачів. У сфері маркетингу трансформери можуть допомогти аналізувати поведінку клієнтів та прогнозувати продажі. Крім того, трансформери можуть бути використані для аналізу мови в різних контекстах, таких як аналіз настрою, виявлення емоцій та аналіз намірів. У статті досліджуються ефективність різних моделей трансформерів на різних рівнях абстракції та в різних контекстах аналізу мови. Проаналізовано ефективність різних моделей мови на різних рівнях абстракції та в різних контекстах аналізу мови.



Підсумки

- Проведено аналіз літератури.
- Проведено аналіз датасетів, які підходять для виконання експерименту.
- Проведено експеримент та порівняльний аналіз результатів.
- Після експерименту можна зазначити, що трансформери є більш ефективним методом ніж класичні методи машинного навчання.

ДОДАТОК Б

Апробація результатів роботи

УДК 004.8
10.30748

DOI:

К.О. Кремененко, І.В. Афанасьєва, К.Г. Онищенко

Харківський національний університет радіоелектроніки, Харків

ДОСЛІДЖЕННЯ МОДЕЛЕЙ НЕЙРОННИХ МЕРЕЖ ТИПУ ТРАНСФОРМЕРИ ДЛЯ СЕМАНТИЧНОГО ТА ЕМОЦІЙНОГО АНАЛІЗУ ПРИРОДНОЇ МОВИ ЛЮДИНИ

У статті досліджуються сучасні моделі нейронних мереж типу трансформери для аналізу природної мови людини, зокрема для семантичного та емоційного аналізу текстів. Розглядаються архітектурні особливості трансформерів, їх переваги перед традиційними методами обробки тексту, а також застосування в задачах розпізнавання емоцій та визначення семантичного значення висловлювань. Проаналізовано ефективність відомих моделей, таких як BERT, RoBERTa, GPT у різних контекстах аналізу текстової інформації. Проведено порівняльний аналіз продуктивності моделей на відкритих наборах даних. Результати дослідження демонструють перспективність використання трансформерів для глибокого розуміння людської мови, що може знайти застосування в чат-ботах, автоматизованих системах підтримки, аналізі соціальних мереж та інших сферах.

Ключові слова: трансформери, семантичний аналіз тексту, емоційний аналіз тексту, BERT, GPT, RoBERTa

Вступ

Постановка проблеми. У сучасному інформаційному просторі, де обсяг текстових даних стрімко зростає, все більш актуальним стає використання автоматизованих методів обробки природної мови (NLP)[1]. Глибинне навчання вже не перший рік відіграє провідну роль у цій галузі, зокрема завдяки появі нових архітектур нейронних мереж. Однією з найбільш впливових інновацій стала архітектура трансформерів[2], що поєднує в собі високий рівень паралелізації та здатність ефективно опрацювати довгі послідовності тексту. Механізм уваги (attention mechanism), який є базовим у трансформерах, уможливує встановлення глибинних зв'язків між окремими словами та фразами без втрати контексту, що раніше було викликом для традиційних моделей типу RNN[3] чи LSTM[4]. Із появою моделей на зразок BERT[5], GPT[6], RoBERTa[7], T5 та їхніх численних варіацій, завдання, пов'язані зі семантичною та емоційною інтерпретацією тексту, стали розв'язуватися значно швидше та точніше. Семантичний аналіз полягає у визначенні змісту, взаємозв'язків між поняттями, розпізнаванні інтонацій автора повідомлення тощо. Водночас

емоційний аналіз (часто відомий як аналіз тональності) спрямований на виявлення та класифікацію емоційного забарвлення, що може включати не лише полярність (позитивний/негативний/нейтральний відтінок), а й конкретні емоції, як-от радість, страх, злість чи сум. Для багатьох прикладних завдань – від автоматизованих систем підтримки клієнтів до соціального моніторингу та аналізу відгуків – здатність моделі правильно інтерпретувати як семантичну структуру, так і емоційні аспекти повідомлень користувачів є надзвичайно важливою. З огляду на це, дослідження трансформерів для семантичного та емоційного аналізу тексту дає змогу значно підвищити ефективність систем, що працюють з людською мовою. Наприклад, в галузі бізнес-аналітики такі рішення можуть бути використані для виявлення ставлення клієнтів до бренду чи продукту, оперативного виявлення негативних чи образливих відгуків, а також надання персоналізованих рекомендацій. У соціальних мережах аналіз емоцій може допомогти ідентифікувати випадки кібербулінгу, пропаганди або маніпулятивних повідомлень. У медичній сфері системи із

вбудованим емоційним аналізом можуть слугувати інструментами психологічної підтримки користувачів або допомагати у виявленні ранніх ознак депресії. З наукового погляду, важливо розуміти, які фактори впливають на результати роботи трансформерів під час аналізу тексту різної природи: від коротких повідомлень у соціальних мережах до великих наукових чи новинних статей. Особливо це стосується мов із розвинутою морфологією (як-от українська), де форми слів можуть значно варіюватися залежно від відмінків, родів і чисел. Тож у центрі даного дослідження перебуває порівняння різних архітектур трансформерів, їхня здатність коректно інтерпретувати семантику та виявляти емоційне забарвлення повідомлень, а також способи підготовки та аугментації даних для досягнення максимальної продуктивності.

Аналіз останніх досліджень і публікацій. У процесі формування теоретичної бази для дослідження моделей нейронних мереж типу трансформери, що застосовуються до семантичного й емоційного аналізу природної мови, були використані наукові статті, монографії, а також матеріали провідних конференцій (ACL, EMNLP, NeurIPS). Критеріями добору джерел слугували такі параметри як: авторитетність, перевага надавалася публікаціям відомих дослідників та груп (Devlin et al., Vaswani et al., Radford et al.), актуальність, достовірність інформації. У праці [8] автори представили базову архітектуру трансформера, побудовану на механізмах самоуваги (self-attention). Цей підхід став фундаментом для всіх наступних моделей на зразок BERT та GPT. Для дослідження ця робота корисна тим, що пояснює теоретичні основи трансформера і демонструє його переваги перед рекурентними нейронними мережами у врахуванні довготривалих залежностей у тексті. У статті [9] автори описують концепцію двонапрявленого попереднього тренування моделі з використанням великих текстових корпусів. Це дало змогу досягти суттєвого поліпшення в різних NLP-завданнях. Робота базується на ідеях BERT, оскільки його глибоке розуміння контексту є ключовим для якісного семантичного та емоційного аналізу. У дослідженні [10] автори презентують великий багатокатегорійний набір даних для емоційного аналізу, що охоплює 27 різних емоцій. Цей ресурс є цінним, оскільки дає змогу тренувати й перевіряти моделі трансформерів не тільки на

базовому негативному/позитивному аналізі, а й на розпізнаванні ширшого спектра емоцій. У роботі [11] команда Google пропонує модель T5, що перетворює всі NLP-завдання на формат "текст у текст". Цей уніфікований підхід полегшує адаптацію моделі до різних задач, включно з тональним та семантичним аналізом. Для дослідження T5 може бути корисною альтернативою або доповненням до BERT-подібних моделей, особливо коли постає питання узагальнення різних типів завдань. У розробці [12] було показано, як покращення гіперпараметрів та розширений тренувальний корпус дають змогу підвищити ефективність BERT. Ця робота ілюструє, що навіть невеликі модифікації та додаткові ресурси для навчання суттєво впливають на результати, що особливо важливо при аналізі тональності та складних емоційних нюансів. У праці [13] пропонується узагальнений авторегресивний підхід до попереднього тренування, який поєднує двонапрявлену обробку тексту з ефективністю авторегресивних моделей. Ця модель демонструє високі показники в задачах класифікації й аналізі тональності.

Мета статті. Основна увага зосереджена на дослідженні моделей нейронних мереж типу трансформери, що застосовуються до семантичного та емоційного аналізу тексту в сучасному інформаційному просторі. Головною метою є виконання ґрунтовних експериментів із цими моделями для визначення найбільш ефективних підходів до аналізу тональності й семантичних зв'язків у природній мові та порівняти їх з класичними методами, такими як SVM[14] та логістична регресія[15]. Для реалізації цього дослідження насамперед потрібен доступ до актуальних і різноманітних корпусів текстів, де кожен приклад анотовано з урахуванням конкретної мети — наприклад, визначеною емоційною категорією чи семантичною роллю. Крім публічно доступних наборів даних, наприклад GoEmotions[16], може виникнути потреба у створенні власної вибірки, зібраної з відкритих джерел або співпраці з партнерами, що надають спеціалізований контент.

Виклад основного матеріалу

Трансформери є одним із найпотужніших рішень у сучасній обробці природної мови, оскільки поєднують у собі здатність паралельної обробки інформації з механізмом самоуваги. Ключовою особливістю такої архітектури стала

відмова від рекурентного підходу, що раніше домінував у нейронних мережах для роботи з послідовностями. Замість зчитування токенів один за одним, як у LSTM чи GRU, трансформери одночасно розглядають усю послідовність, аналізуючи зв'язки між усіма словами або символами через багатосаровий механізм уваги. Завдяки цьому модель позбавлена обмежень, пов'язаних із поступовим «забуванням» контексту й накопиченням похибок, а також здатна фокусуватися на віддалених взаємозв'язках, які залишалися малодоступними для попередніх архітектур. Така паралельна схема не лише збільшує пропускну здатність у процесі навчання, а й поліпшує розпізнавання нюансів мови, оскільки кожне слово оцінюється з огляду на весь контекст відразу.

У семантичному аналізі трансформери дають змогу глибше досліджувати лексичні та синтаксичні зв'язки між словами, визначати, до яких концептів вони належать і як утворюють складні структури значень. Така здатність виникає завдяки багатоголовій увазі, де кожна «голова» може виявляти різні патерни взаємодії між словами. Це дозволяє з високою точністю витягувати сутності, розпізнавати референтні зв'язки чи відновлювати пропущені елементи речення. Крім того, трансформери природним чином долають проблему зникаючих і вибухових градієнтів, характерну для рекурентних мереж, оскільки не покладаються на довге передавання станів від одного кроку до іншого. Замість цього вони використовують резидульні з'єднання та нормалізаційні шари, що стабілізують процес навчання й дають змогу масштабувати моделі на величезні обсяги даних. В емоційному аналізі механізм самоуваги відкриває додаткові можливості для виявлення тонких відтінків настрою. Якщо класичні підходи здебільшого покладалися на специфічні маркери (наприклад, «щасливий», «злий») і лексичні словники, то трансформери можуть враховувати цілі фрази та навіть контекст, розподілений по кількох реченнях. Це дає змогу точніше визначати емоційний зміст у випадках сарказму чи іронії, де пряме значення слів не відповідає справжній тональності. Завдяки паралельній обробці весь набір слів оцінюється миттєво, й модель «помічає» не лише формальні негативно забарвлені лексеми, а й супутні сигнали, що вказують на прихований або завуальований намір мовця. Усе це робить трансформери одним із основних

інструментів у задачах емоційного аналізу, полегшуючи розуміння психологічних чи соціальних аспектів тексту і відкриваючи можливості для автоматизації складних процесів модерації, персоналізованого маркетингу чи навіть побудови систем емоційно орієнтованої взаємодії.

Для виконання запланованого дослідження розглядається декілька моделей на базі архітектури трансформерів, серед яких передусім виділяються BERT, RoBERTa та GPT. Кожна з них пропонує унікальний підхід до попереднього тренування і має власні характеристики, що можуть виявитися визначальними в умовах аналізу семантики й емоційного забарвлення тексту.

Першою моделлю, яка справила суттєвий вплив на всю галузь, є BERT (Bidirectional Encoder Representations from Transformers). Вона ґрунтується на двонапрявленому опрацюванні контексту, що означає можливість «читати» речення одночасно зліва направо й справа наліво. Завдяки цій особливості BERT краще засвоює цілісний зміст, де значення конкретного слова чи фрази великою мірою залежить від їхнього оточення. На стадії попереднього тренування модель виконує масковане мовне моделювання, коли певна частина токенів у реченні замінюється на спеціальний символ [MASK], а завдання полягає в тому, щоб правильно передбачити приховані слова. Це сприяє формуванню детального уявлення про лексичні та синтаксичні структури мови. Крім того, оригінальна версія BERT містила компонент Next Sentence Prediction, завдяки якому модель училася визначати, чи логічно одне речення продовжує інше. Хоча згодом цей етап був підданий сумніву щодо його ефективності, він все ж вплинув на формування загальної тренувальної методології.

Модель RoBERTa (Robustly Optimized BERT Approach) є результатом масштабного доопрацювання початкової версії BERT. Дослідники збільшили обсяг даних для попереднього тренування і відмовилися від задачі Next Sentence Prediction, зосередившись на якіснішому маскуванні й досконалішому налаштуванні гіперпараметрів. У RoBERTa змінено й сам підхід до маскування: якщо BERT використовував статичне маскування (коли всі заміни токенів визначалися один раз), то в RoBERTa воно динамічне, що дає змогу під час

кожного проходу за новою епохою змінювати набір замаскованих слів. Така стратегія уможливіє кращу генералізацію, адже модель зустрічається з різними варіантами пропущених слів. Розробники також використали суттєво більші обчислювальні ресурси, тренуючи модель протягом триваліших періодів і на ширших корпусах, що підвищило кінцеву ефективність у низці задач, включаючи аналіз тональності та розпізнавання сутностей.

Інший напрям дослідження утілено в серії моделей GPT. Їхній генеративний характер дає змогу передбачати наступне слово на основі всіх попередніх, що робить GPT особливо цікавим для аналізу діалогів чи будь-яких текстів, де важливо зрозуміти логіку продовження висловлювання. GPT корисний також у завданнях, де є потреба виявляти загальний настрій або передбачати відповідну реакцію на заданий контекст. Головна відмінність від BERT-орієнтованих підходів полягає в тому, що GPT здатний краще відтворювати послідовність мовлення, тоді як BERT основну увагу приділяє двонапрявленому розгляду внутрішньої структури речення.

Окремо варто згадати й про інші розробки, тісно пов'язані з цими базовими моделями. Наприклад, DistilBERT[17] створений для зменшення обсягу параметрів і пришвидшення часу обчислень без суттєвих втрат у точності. Такий підхід використовує техніку дистилляції знань: «легша» модель успадковує репрезентації від «важчої» базової. XLNet[18], зі свого боку, поєднує ідеї автогресивних методів (як у GPT) із двонапрявленим контекстом, близьким до BERT, тим самим прагнучи використати переваги обох підходів. Застосування цих варіацій залежить від конкретних вимог проекту: якщо домінує потреба в економії ресурсів, доречніше DistilBERT, якщо ж важлива максимально можлива точність у складних лінгвістичних сценаріях — варто зупинитися на RoBERTa або навіть GPT-3.5[19], що є послідовником концепцій GPT з істотно більшим числом параметрів.

У межах дослідження передбачається оцінити, як саме відрізнятимуться результати під час виконання задач семантичного й емоційного аналізу залежно від обраної архітектури: BERT, RoBERTa чи GPT. На одному корпусі текстів ці моделі проходять однакові випробування за спільною методикою, що дасть змогу з'ясувати, чи виправдана складність і ресурсомісткість більш досконалих варіантів на кшталт RoBERTa

порівняно зі спрощеними (наприклад, DistilBERT). Особливий акцент робитиметься на тому, наскільки глибоко модель враховує контекст, уникає розповсюджених помилок під час обробки полісемії чи складних лексико-семантичних конструкцій і як добре вона визначає емоційні відтінки, включаючи сарказм чи іронію. Це порівняння буде найбільш інформативним, якщо вмістить широке коло експериментальних сценаріїв, у яких одночасно оцінять здатність моделі реагувати на довжину речень, різновиди синтаксичних структур і різномірність словникового запасу.

У процесі дослідження заплановано серію експериментів, метою яких є визначити, наскільки точно та надійно обрані трансформерні моделі (BERT, RoBERTa, GPT) здатні виконувати семантичний і емоційний аналіз тексту. Основними метриками в цьому контексті залишаються показники точності, повноти й F1-міри, оскільки вони належать до найбільш розповсюджених критеріїв для задач класифікації. При цьому важливо розуміти, як ці метрики обчислюються і які їхні обмеження.

Точність (precision) відображає частку правильно спрогнозованих позитивних результатів серед усіх, які модель визнала позитивними. Зазвичай цю метрику використовують тоді, коли критично мінімізувати хибні тривоги. Визначення можна виразити через кількість істинно позитивних передбачень TP і помилково позитивних передбачень FP.

Формула набуває вигляду (1):

$$Precision = \frac{TP}{TP + FP} \cdot (1)$$

Повнота (recall) фіксує частку правильних позитивних прогнозів серед усіх реальних позитивних випадків, наявних у тестовому наборі. Цей показник особливо важливий тоді, коли небажано пропускати потенційно важливі (наприклад, токсичні чи негативні) приклади. Якщо FN означає кількість помилково негативних передбачень, тоді повнота обчислюється за формулою (2):

$$Recall = \frac{TP}{TP + FN} \cdot (2)$$

Для досягнення балансу між точністю й повнотою зазвичай звертаються до F1-міри, що є гармонійним середнім цих двох показників. Обчислення F1-міри дає більш зважену оцінку

загальної ефективності моделі, коли важливо одночасно враховувати кількість помилкових тривог і «пропусків». Формула для F1-міри має вигляд (3):

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \cdot (3)$$

Залежно від розподілу даних у корпусі можуть виникати ситуації, коли одна чи кілька емоційних або семантичних категорій зустрічаються набагато частіше, ніж інші. У такому випадку постає необхідність у використанні підходів до усереднення цих показників.

Зокрема, *micro*-усереднення (*micro*-F1) розглядає загальну суму TP, FP та FN для всіх класів, обчислюючи метрики як єдине ціле. Цей метод підходить, коли важить загальна частка правильних рішень без урахування того, до якого класу вони належать.

Натомість *macro*-усереднення (*macro*-F1) передбачає обчислення точності, повноти й F1-міри для кожного класу окремо, а потім усереднення цих значень. Така стратегія дозволяє зрозуміти, наскільки добре модель працює з рідкісними або слабо представленими емоціями й семантичними відтінками тексту. Якщо ж один із класів (наприклад, «нейтральний») переважає, модель може загалом демонструвати хороші результати з погляду звичайної F1-міри, але водночас погано розрізняти менш поширені категорії, що нерідко трапляється в реальних ситуаціях.

Окрім зазначених метрик, результативність можна додатково оцінити через детальний аналіз помилок, тобто перегляд прикладів, у яких модель невірно класифікувала емоційне забарвлення чи допустилася непослідовності в семантичних висновках. Такий підхід допоможе визначити «болючі точки» системи, наприклад, випадки сарказму або тексти зі складною структурою речень, де зв'язки між словами неявні. Крім того, варто враховувати конфузійну матрицю, що відображає, як модель плутає класи між собою. Якщо модель часто трактує «радість» як «нейтральність» або навпаки, це може свідчити про неточності в анотації даних чи потребу в доопрацюванні способів донавчання.

Зрештою, такий комплексний підхід до оцінювання — від базових метрик до розгорнутого аналізу помилок — дозволить отримати вичерпне уявлення про те, наскільки обрані трансформерні моделі справляються зі складними завданнями

семантичного й емоційного аналізу, та допоможе сформулювати конкретні рекомендації щодо подальшого поліпшення системи.

Слід згадати про класичні методи машинного навчання, тому що в ході експерименту ми будемо порівнювати їх з трансформерами. Перший метод, який буде порівнюватися — SVM. Це потужний алгоритм машинного навчання, що використовується для задач класифікації, регресії та навіть виявлення аномалій. Його основна ідея полягає в тому, щоб знайти оптимальну гіперплощину, яка найкраще розділяє дані двох класів, залишаючи між ними максимально можливий зазор. Цей метод особливо добре працює на малих і середніх наборах даних, де класи можна чітко розділити.

SVM працює за принципом пошуку найбільш віддалених точок від гіперплощини розділення, які називаються опорними векторами. Саме ці точки визначають положення гіперплощини і її орієнтацію в просторі. В інших алгоритмах класифікації, таких як логістична регресія, використовується підхід мінімізації помилок, тоді як у SVM головна мета — максимізувати відстань між класами, що підвищує стійкість моделі до шуму в даних.

Другим методом є метод логістичної регресії. Це статистичний метод машинного навчання, який використовується для задач класифікації, зокрема для бінарної класифікації, де метою є передбачення ймовірності належності об'єкта до одного з двох класів. Попри свою назву, логістична регресія не є типом регресії в класичному розумінні, оскільки її мета — передбачити клас, а не безперервну змінну.

Основною ідеєю логістичної регресії є використання логістичної функції, яка перетворює лінійну комбінацію вхідних ознак в ймовірність, яка лежить у межах від 0 до 1.

Для виконання дослідження було обрано два корпуси, орієнтовані на різні аспекти аналізу тексту. Першим став набір GoEmotions. є анотований датасет для аналізу емоцій у тексті. Його основна мета — надати великий набір текстових даних для глибокого вивчення емоційних реакцій у природному мовленні. Датасет містить приблизно 58 000 текстових уривків, узятих із коментарів на різних веб-платформах.

Ключовою особливістю GoEmotions є його детальна емоційна розмітка. Датасет містить 27

різних емоційних міток, що дає змогу класифікувати текстові уривки набагато точніше, ніж традиційні підходи з бінарною (позитивною або негативною) чи спрощеною емоційною розміткою. Серед представлених емоцій є такі, як радість, смуток, злість, вдячність, розчарування, страх, здивування, задоволення та інші. Окрім того, є окрема категорія для нейтральних коментарів, які не виражають жодних особливих емоцій.

Анотація даних у GoEmotions проводилася ретельно, із залученням кількох незалежних анотацій для кожного уривка. Це допомогло зменшити суб'єктивність оцінки та підвищити надійність емоційної класифікації. Кожен коментар міг отримати кілька міток, що відображає складність та багатовимірність людських емоцій. Наприклад, одна й та сама фраза може одночасно виражати як злість, так і розчарування або, навпаки, вдячність та радість. Другим набором став IMDb Large Movie Review Dataset[20]. Це великий корпус текстових рецензій із сайту IMDb, розроблений для дослідження аналізу тональності тексту. Його створили дослідники з університету Стенфорда. Цей датасет став стандартним тестовим набором для багатьох досліджень у галузі аналізу настрою та глибокого навчання.

До складу набору входять 50 000 рецензій на фільми, поділені на дві рівні частини: 25 000 для тренувальної вибірки та 25 000 для тестування. Всі рецензії є розлогими текстами, у яких автори детально обґрунтовують свою думку щодо переглянутого фільму. Датасет містить два основні класи настрою: позитивний та негативний, при цьому рецензії з оцінками 5 і 6 (за десятибальною шкалою IMDb) були вилучені, щоб уникнути нейтральних відгуків. Таким чином, лише тексти з рейтингами 7–10 вважаються позитивними, а 1–4 — негативними.

Однією з важливих особливостей IMDb Large Movie Review Dataset є те, що він не містить дублікатів рецензій у тренувальній та тестовій вибірках. Це гарантує, що модель не запам'ятає конкретні відгуки і справді навчатиметься узагальнювати патерни позитивного й негативного мовлення. Датасет також включає 50 000 додаткових неанотованих рецензій, які можуть використовуватися для неконтрольованого навчання.

Цей датасет вже доказав свою корисність і є актуальним, тому що зараз його вже застосовують

у побудові систем автоматичної модерації коментарів, розробці чат-ботів, дослідженнях обробки природної мови та емоційної аналітики. Через свою структурованість та якість.

Завдяки цьому матеріалу можна дослідити, як трансформери справляються з визначенням загального настрою автора, коли текст не обмежується короткими висловлюваннями, а розгортається на кілька абзаців. Обидва корпуси містять досить розмаїтий лексичний склад, що допомагає перевірити універсальність обраних моделей і їхню чутливість до різних контекстів.

На етапі підготовки корпуси були оброблені з урахуванням типових вимог до навчання трансформерів. Спочатку з текстів видаляли зайві символи, що не несуть змістовного навантаження: HTML-теги, спеціальні розмітки чи емодзі, які могли б викривляти лінгвістичну статистику. Далі здійснювалася токенизація, узгоджена з вимогами кожної моделі. Для BERT і RoBERTa використовувався підхід WordPiece[21], тоді як GPT базується на Byte-Pair Encoding[22], однак загальна логіка залишається подібною: текст перетворюється на послідовність підслівників так, щоб відобразити морфологічне різноманіття. Після завершення токенизації в текст додавалися спеціальні маркери (CLS або SEP), якщо це передбачено архітектурою. Корпуси були розділені на навчальну, валідаційну та тестову вибірку в приблизному співвідношенні 80%, 10% і 10%. Така структура дає змогу проводити контроль якості на проміжних етапах і запобігає перенавчанню, оскільки результати на валідаційній вибірці слугують орієнтиром для коригування гіперпараметрів. Результати експерименту будуть відображені в таблиці 1.

Таблиця 1
Результати першого експерименту

Модель	Набір даних	Точність	Повнота	F1-міра
Логістична регресія	GoE	67.2%	63.1%	65.0%
Логістична регресія	IMDb	72.4%	69.8%	71.1%
SVM	GoE	70.4%	66.7%	68.5%

Продовження таблиці 1

SVM	IMDb	75.8%	72.9%	74.3 %
BERT	GoE	83.1%	80.2%	81.6 %
BERT	IMDb	80.5%	76.8%	78.6 %
RoBERTa	GoE	85.0%	82.3%	83.6 %
RoBERTa	IMDb	82.8%	79.4%	81.0 %
GPT	GoE	80.5%	78.0%	79.2 %
GPT	IMDb	78.5%	75.1%	76.8 %

Джерело: розроблено авторами.

Результати першого експерименту показали, що трансформери значно перевершують класичні методи, такі як логістична регресія та SVM, у задачах класифікації тексту. Найкращі показники продемонструвала модель RoBERTa, яка досягла найвищої точності як на GoEmotions, так і на IMDb, що підтверджує її здатність ефективно враховувати контекст та тонкі мовні нюанси. GPT також показала хороші результати, особливо на наборі IMDb, де текстові дані є більш довгими та розгорнутими.

Після проведення першого етапу експерименту проаналізовано результати та виявлено кілька аспектів, які можна покращити для підвищення точності моделей. Основні зміни торкнулися балансування класів, оптимізації навчального процесу, розширення даних та роботи з довгими текстами.

По-перше, звернено увагу на проблему дисбалансу класів у наборі GoEmotions, де деякі емоції зустрічалися значно рідше за інші. Для цього було впроваджено функцію втрат Focal Loss, яка зменшує вагу добре класифікованих прикладів і змушує модель приділяти більше уваги менш поширеним категоріям. У наборі IMDb, де частка позитивних рецензій була дещо вищою, застосували стратифіковане розбиття, щоб забезпечити рівномірний розподіл класів у навчальній, валідаційній та тестовій вибірках.

Ще одне суттєве покращення — розширення даних. Оскільки в GoEmotions містяться короткі повідомлення, застосовано Easy Data Augmentation (EDA), яка включала перефразування, заміну слів синонімами та випадкове видалення малозначущих слів. Це дозволило моделі краще розпізнавати варіативність тексту. У випадку IMDb, де тексти

значно довші, було використано Back Translation: рецензії спочатку перекладалися іншими мовами (наприклад, англійська → німецька → англійська), а потім поверталися назад, що створювало нові варіанти записів без втрати смислу.

Останнім покращенням стало ефективніше опрацювання довгих текстів. Для IMDb було реалізовано Sliding Window Attention, що дозволило моделям BERT і RoBERTa обробляти довші рецензії, розбиваючи їх на частини із збереженням контексту. Для GPT було збільшено максимальну довжину контексту, що дало змогу краще аналізувати відгуки, що містять кілька абзаців.

Ці покращення дозволили зробити моделі більш стійкими до різноманітності текстів, зменшити проблему дисбалансу та покращити загальну якість класифікації.

Отже, результати експерименту після покращення будуть відображені в таблиці 2.

Таблиця 2

Результати другого експерименту

Модель	Набір даних	Точність	Повнота	F1-міра
Логістична регресія	GoE	68.9%	65.0%	66.9 %
Логістична регресія	IMDb	74.1%	71.2%	72.6 %
SVM	GoE	71.8%	68.5%	70.1 %
SVM	IMDb	77.5%	74.8%	76.1 %
BERT	GoE	86.0%	83.7%	84.8 %
BERT	IMDb	83.2%	80.1%	81.6 %
RoBERTa	GoE	88.1%	85.6%	86.8 %
RoBERTa	IMDb	85.9%	82.9%	84.4 %
GPT	GoE	82.8%	80.5%	81.6 %
GPT	IMDb	81.3%	78.6%	79.9 %

Джерело: розроблено авторами.

Завдяки цим покращенням вдалося підвищити точність трансформерів на 2-4%, а F1-міру для рідкісних класів у GoEmotions – аж на 6%. Окрім цього, стабілізувалося навчання моделей, а точність SVM та логістичної регресії також зросла, хоч і не так суттєво, як у нейромережових підходів.

Висновки

Проведене дослідження підтверджує, що архітектури трансформерів, такі як BERT, RoBERTa та GPT, суттєво підвищують точність і надійність систем семантичного та емоційного аналізу порівняно з традиційними методами (логістична регресія, SVM). Механізм самоуваги дозволяє краще враховувати контекст і складні лінгвістичні залежності в тексті, що особливо важливо для розпізнавання тонких відтінків емоцій і смислових зв'язків між словами.

Наведені експерименти свідчать про те, що попередньо навчені моделі ефективно переносять глобальні мовні знання, набуті під час тренування на великих корпусах, у конкретні прикладні завдання на кшталт класифікації позитивних чи негативних відгуків або визначення спектра емоцій у повідомленнях користувачів. Завдяки адаптації (до цільових наборів даних можна істотно підвищити продуктивність навіть при обмеженій кількості анованих прикладів, особливо якщо збалансувати та ретельно підготувати навчальну й тестову вибірки. У дослідженні також продемонстровано, що різні

трансформери можуть давати трохи відмінні результати: RoBERTa в ряді випадків випереджає базову версію BERT, тоді як GPT виявляється особливо корисною для завдань, пов'язаних із генеруванням тексту або передбаченням продовження думки.

Усе це підкреслює перспективність трансформерів як платформи для розбудови складніших систем аналізу тексту, включно з багатокомпонентними інструментами моніторингу й модерації контенту. Водночас результати вказують і на низку обмежень: обчислювальні ресурси, необхідні для тренування моделей, можуть бути значними, а пояснюваність прийнятих рішень усе ще залишається складним питанням. Подальші дослідження можуть спрямовуватися на оптимізацію архітектур (через праунінг, квантування, знеглиблення моделей), а також на глибший аналіз культурних і лінгвістичних особливостей, що набуває особливого значення в задачах багатомовного емоційного аналізу. Загалом, методи трансформерів суттєво покращують можливості семантичного й емоційного розуміння тексту, закладаючи підґрунтя для нових, більш інтелектуальних рішень у галузі обробки природної мови.

Список літератури

1. Що таке обробка природної мови (NLP) [Електронний ресурс]. – URL: <https://metinvest.digital/ru/page/1052> (дата звернення: 21.12.2024).
2. What is transformer model? [Електронний ресурс]. – URL: <https://www.ibm.com/think/topics/transformer-model> (дата звернення: 22.12.2024)
3. Назаренко Д.С., Афанасьєва І.В., Голян Н.В. Нейромережевий підхід для емоційного розпізнавання тексту. Біоніка інтелекту. 2019. Т. 1, № 92. С. 9–14.
4. Назаренко Д.С., Афанасьєва І.В., Голян Н.В. Investigation of the Deep Learning Approaches to Classify Emotions in Texts // CEUR Workshop Proceedings, 2021, 2870, С. 206-224.
5. BERT [Електронний ресурс] – URL: <https://h2o.ai/wiki/bert/> (дата звернення: 30.12.2024).
6. What is GPT (generative pretrained transformer) [Електронний ресурс]– URL: <https://www.ibm.com/think/topics/gpt> (дата звернення: 30.12.2024).
7. Introducing RoBERTa Base Model: A Comprehensive Overview [Електронний ресурс] – URL: https://medium.com/@marketing_novita.ai/introducing-roberta-base-model-a-comprehensive-overview-330338afa082 (дата звернення: 30.12.2024).
8. Attention Is All You Need [Електронний ресурс] – URL: <https://arxiv.org/abs/1706.03762> (дата звернення: 02.01.2025).
9. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [Електронний ресурс] – URL: <https://arxiv.org/abs/1810.04805> (дата звернення: 02.01.2025).
10. GoEmotions: A Dataset of Fine-Grained Emotions [Електронний ресурс] – URL: <https://arxiv.org/abs/2005.00547> (дата звернення: 26.12.2024).
11. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer [Електронний ресурс] – URL: <https://arxiv.org/abs/1910.10683> (дата звернення: 26.12.2024).
12. RoBERTa: A Robustly Optimized BERT Pretraining Approach [Електронний ресурс] – URL: <https://arxiv.org/abs/1907.11692> (дата звернення: 30.01.2025).

13. XLNet: Generalized Autoregressive Pretraining for Language Understanding [Електронний ресурс] – URL: <https://arxiv.org/abs/1906.08237> (дата звернення: 02.02.2025).
14. What are support vector machines (SVMs) [Електронний ресурс] – URL: [https://www.ibm.com/think/topics/support-vector-machine#:~:text=A%20support%20vector%20machine%20\(SVM,the%201990s%20by%20Vladimir%20N.](https://www.ibm.com/think/topics/support-vector-machine#:~:text=A%20support%20vector%20machine%20(SVM,the%201990s%20by%20Vladimir%20N.) (дата звернення: 04.02.2025).
15. What is logistic regression? [Електронний ресурс] – URL: <https://www.ibm.com/think/topics/logistic-regression> (дата звернення: 05.02.2025).
16. GoEmotions [Електронний ресурс] – URL: <https://www.kaggle.com/datasets/debarshichanda/goemotions> (дата звернення: 06.02.2025).
17. DistilBERT: A Smaller, Faster, and Distilled BERT [Електронний ресурс] – URL: <https://zilliz.com/learn/distilbert-distilled-version-of-bert> (дата звернення: 10.02.2025).
18. XLNet [Електронний ресурс] – URL: https://huggingface.co/docs/transformers/model_doc/xlnet
19. OpenAI GPT-3.5 [Електронний ресурс] – URL: <https://lablab.ai/tech/openai/gpt3-5> (дата звернення: 12.02.2025).
20. IMDB Dataset of 50k Movie Reviews [Електронний ресурс] – URL: <https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>
21. What is WordPiece [Електронний ресурс] – URL: <https://h2o.ai/wiki/wordpiece/> (дата звернення: 15.02.2025).
22. Byte-Pair Encoding tokenizations [Електронний ресурс] – URL: <https://huggingface.co/learn/nlp-course/chapter6/5> (дата звернення: 17.02.2025).

Відомості про авторів:**Кремененко Кирило Олегович**

студент

Харківського національного університету
радіоелектроніки,
Харків, Україна<https://orcid.org/0009-0002-2050-660X>**Information about the authors:****Kremenenko Kyrylo Olegovych**

student

Kharkiv national university of radioelectronics,
Kharkiv, Ukraine<https://orcid.org/0009-0002-2050-660X>**Афанасьєва Ірина Віталіївна**

доцент

кандидат технічних наук

Харківського національного університету
радіоелектроніки,
Харків, Україна<https://orcid.org/0000-0003-4061-0332>**Afanasieva Iryna Vitaliivna**

docent

candidate of technical sciences

Kharkiv national university of radioelectronics,
Kharkiv, Ukraine<https://orcid.org/0000-0003-4061-0332>**Онищенко Костянтин Георгійович**

старший викладач

Харківського національного університету
радіоелектроніки,
Харків, Україна<https://orcid.org/0000-0002-7746-4570>**Onyshchenko Kostiantyn Heorhiiovych**

senior teacher

Kharkiv national university of radioelectronics,
Kharkiv, Ukraine<https://orcid.org/0000-0002-7746-4570>**RESEARCH OF TRANSFORMER-TYPE NEURAL NETWORK MODELS FOR SEMANTIC AND EMOTIONAL ANALYSIS OF HUMAN NATURAL LANGUAGE**

K. Kremenenko, I. Afanasieva, K. Onyshchenko

The article explores modern transformer-based neural network models for analyzing human natural language, with a particular focus on semantic and emotional text analysis. It examines the architectural features of transformers, their advantages over traditional text processing methods, and their applications in tasks such as emotion recognition and semantic meaning extraction.

The study analyzes the effectiveness of well-known models, including BERT, RoBERTa, and GPT, in various text analysis contexts. A comparative performance evaluation of these models is conducted using open datasets to assess their accuracy and efficiency. The results highlight the potential of transformers for deep language understanding, demonstrating their applicability in chatbot development, automated support systems, social media analysis, and other fields requiring advanced natural language processing capabilities.

Additionally, the research discusses the challenges associated with transformer models, such as computational requirements and data biases, while also considering possible optimizations and future advancements in the field. The findings suggest that transformer-based approaches will continue to play a crucial role in enhancing the accuracy and contextual awareness of AI-driven language processing systems.

Keywords: transformers, semantic text analysis, emotional text analysis, BERT, GPT, RoBERTa

ДОВІДКА

Наукова стаття авторів: К.О. Кремененко, І.В. Афанасьєва, К.Г. Онищенко “Дослідження моделей нейронних мереж типу трансформери для семантичного та емоційного аналізу природньої мови людини” (обсяг статті – 8 сторінок формату А4) позитивно розглянута на засіданні редакційної колегії наукового фахового видання “Збірник наукових праць Харківського національного університету Повітряних Сил” та схвалена до друку у випуску №2(84) 2025 року.

Друк заплановано на кінець вересня 2025 року.

“Збірник наукових праць Харківського національного університету Повітряних Сил ” занесений до категорії “Б” Переліку наукових фахових видань України (технічні та військові науки за спеціальностями 122, 123, 126, 172, 253, 255, 272, 275), затверджено наказами Міністерства освіти і науки України від 17.03.2020 № 409 та від 02.07.2020 № 886, включено до міжнародної наукометричної бази даних Index Copernicus (Польща).

Відповідальний за випуск наукових фахових видань
Харківського національного університету Повітряних Сил
імені Івана Кожедуба
доктор філософії



Вадим ФУСТІЙ

ДОДАТОК В

Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ




Дата звіту 6/11/2025
Дата редагування ---



Звіт не був оцінений

Звіт подібності

метадані

Назва організації
Kharkiv National University of Radio Electronics

Заголовок
Звіт_2025_M_ПІ_ПЗм-23-2_Кремененко_К_О_ скорочений

Автор Науковий керівник / Експерт
Кремененко Кирилл ОлеговичАфанасьєва І.В.Нечволод В.Ю.

підрозділ
каф. ПІ

Обсяг знайдених подібностей

Коефіцієнт подібності вивчає, який відсоток тексту по відношенню до загального обсягу тексту було знайдено в різних джерелах. Зверніть увагу, що високі значення коефіцієнта не автоматично означають плагіат. Звіт має аналізувати компетентна / уповноважена особа.

4.83%
4.83%

КП 1

25

Довжина фраз для коефіцієнта подібності 2

1.17%
1.17%

КЦ

7787

Кількість слів


1.17%
1.17%

КЦ

62274

Кількість слів/слів

3	2_стаття_кремененю_2 4/8/2025 Kharkiv National University of Air Force named after Ivan Kozhedub (Kharkiv National University of Air Force named after Ivan Kozhedub)	41 0.53 %
4	https://openarchive.nure.ua/bitstreams/fe1558bf-4137-4489-884d-3f977293b8d6/download	15 0.19 %
5	Analysis of Sentiment Towards Starbucks on Twitter 6/24/2024 Universiteit Gent (Plagiatcontrole-tool)	14 0.18 %
6	Синицька_бакалаврська_2024_керівник Робеєко 6/5/2024 Taras Shevchenko National University of Kyiv (Навчально-науковий інститут філології)	14 0.18 %
7	https://www.cs.utexas.edu/~gdumett/courses/online-course/materials.html	13 0.17 %
8	https://www.cs.utexas.edu/~gdumett/courses/online-course/materials.html	13 0.17 %
9	Синицька_бакалаврська_2024_керівник Робеєко 6/5/2024 Taras Shevchenko National University of Kyiv (Навчально-науковий інститут філології)	13 0.17 %
10	https://elar.tntu.edu.ua/bitstream/5b454621-%D0%A1%D0%BE%D1%80%D0%BE%D1%86%D1%8C%D0%BA%D0%B8%D0%B9%20%D0%A1%D0%AZ%D1%81-41%20%D0%B1%D0%B0%D0%BA%D0%B0%D0%BB%D0%B0%D0%B2%D1%80%202024.pdf	12 0.15 %
з бази даних RefBooks (0.00 %)		
ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВОК	КІЛЬКІСТЬ ІДЕНТИФІКАЦІЙНИХ СЛІВ (ФРАГМЕНТІВ)
з домашньої бази даних (0.00 %)		
ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВОК	КІЛЬКІСТЬ ІДЕНТИФІКАЦІЙНИХ СЛІВ (ФРАГМЕНТІВ)
з програми обміну базами даних (3.30 %)		
ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВОК	КІЛЬКІСТЬ ІДЕНТИФІКАЦІЙНИХ СЛІВ (ФРАГМЕНТІВ)
1	2_стаття_кремененю_2 4/8/2025 Kharkiv National University of Air Force named after Ivan Kozhedub (Kharkiv National University of Air Force named after Ivan Kozhedub)	155 (5) 1.99 %
2	Синицька_бакалаврська_2024_керівник Робеєко 6/5/2024 Taras Shevchenko National University of Kyiv (Навчально-науковий інститут філології)	38 (4) 0.49 %
3	Analysis of Sentiment Towards Starbucks on Twitter 6/24/2024 Universiteit Gent (Plagiatcontrole-tool)	34 (4) 0.44 %
4	АКТ-2023М/27КТ.pdf 12/25/2023 Ukrainian Academy of Printing (Кафедра АКТ)	13 (2) 0.17 %
5	Диплом Зиморой В В.docx 12/18/2023 Sumy State University (Кафедра наноелектроніки)	7 (1) 0.09 %
6	2024_Maгістерська_ІПм_23_3_Соколов_В_І 12/14/2024 Ivano-Frankivsk National Technical University of Oil and Gas (Каф. ІПЗ Магістр)	5 (1) 0.06 %

7	Прогнозування економічних процесів на основі новин з відкритих джерел за допомогою нейронних мереж 3/16/2025 National Technical University of Ukraine Igor Sikorskyi Kyiv Politech Institute (National Technical University of Ukraine Igor Sikorskyi Kyiv Politech Institute)	5 (1) 0.06 %
з Інтернету (1.53 %)		
ПОРЯДКОВИЙ НОМЕР	ДЖЕРЕЛО URL	КІЛЬКІСТЬ ОДНАКОВИХ СЛІВ (ФРАГМЕНТІВ)
1	https://www.cs.utexas.edu/~gdumett/courses/online-course/materials.html	44 (4) 0.57 %
2	https://elartu.tntu.edu.ua/bitstream/handle/45462/1/%D0%A1%D0%BE%D1%80%D0%BE%D1%86%D1%8C%D0%BA%D0%BB%D0%89%20%D0%A1%D0%A2%D1%81-41%20%D0%B1%D0%B0%D0%8A%D0%B0%D0%BB%D0%B0%D0%B2%D1%80%202024.pdf	34 (3) 0.44 %
3	https://openarchive.nure.ua/bitstreams/ite1558bf-4137-4489-884d-3f977299b8a5/download	26 (2) 0.33 %
4	https://blog.cadn.net/u013172930/article/details/147232344	8 (1) 0.10 %
5	https://cloud.tencent.com/developer/article/2469126	7 (1) 0.09 %
Список прийнятих фрагментів (немає прийнятих фрагментів)		
ПОРЯДКОВИЙ НОМЕР	ЗМІСТ	КІЛЬКІСТЬ ОДНАКОВИХ СЛІВ (ФРАГМЕНТІВ)

ДОДАТОК Г

Експертний висновок результатів перевірки кваліфікаційної роботи на
відповідність оформлення вимогам ДСТУ 3008: 2015

Експертний висновок результатів перевірки кваліфікаційної роботи

студент
(посада)

програмної інженерії
(кафедра)

ІІЗм-23-2□
(група)

Кирило КРЕМЕНЕНКО

(прізвище, ім'я, по батькові)

Зауваження

Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	7.1 Загальні положення	
	7.3 Нумерація сторінок звіту	
	7.5 Рисунки	
	7.6 Таблиці	
	7.7 Переліки	
	7.8 Примітки	
	7.9 Виноски	
	7.10 Формули та рівняння	
	7.11 Посилання	
	7.13 Список авторів	
	7.14 Скорочення та умовні позначки	
	7.15 Додатки	

Експерт

Зауважень немає.
15.06.2025

Вадим НЕЧВОЛОД

(прізвище, ініціали)