

РЕЦЕНЗІЯ

рецензента, кандидата технічних наук, доцента

Філімончук Тетяни Володимирівни

на дисертаційну роботу Гаврашенка Антона Олеговича

на тему «**Модель та методи обробки текстових масивів в інформаційній системі перекладу штучних мов**»,

подану до захисту на здобуття наукового ступеня доктора філософії

з галузі знань 12 Інформаційні технології

спеціальності 126 Інформаційні системи та технології

Актуальність теми дослідження.

Розглянута тематика є важливою та актуальною в умовах зростання обсягів мультимодальної інформації. Удосконалення методів обробки даних різних модальностей із збереженням точності та продуктивності є важливим, у тому числі для розвитку комунікаційних платформ, де необхідна безпечна, масштабована та персоналізована взаємодія між учасниками. Забезпечення автентичності, адаптивної роботи різними мовами таких систем є ключовим чинником у створенні надійної цифрової інфраструктури освіти й науки, що підтверджує високий науковий і прикладний потенціал дослідження.

Ступінь обґрунтованості наукових положень.

Наукові положення, викладені в дисертаційній роботі, мають належний ступінь обґрунтованості та підтверджуються комплексним використанням системного аналізу, методів математичного моделювання, алгоритмів глибинного навчання та експериментальних досліджень. Запропонована модель розподіленої інформаційної системи та вдосконалені методи обробки мультимодального інформаційного контенту побудовані на основі теоретичних досліджень і підтверджені чисельними розрахунками та моделюванням.

У дисертації проведено ґрунтовний аналітичний огляд предметної

області з детальним аналізом сучасних підходів до обробки текстових, голосових та візуальних даних у системах дистанційної комунікації. Автором чітко визначено мету, об'єкт, предмет та завдання дослідження, які узгоджуються з отриманими результатами.

У другому **розділі** було досліджено статистичний модуль, що реалізує генерацію словників на основі паралельних текстів із застосуванням алгоритмів двостороннього покращення, модифікованого алгоритму Кнута-Морріса-Пратта для подібних мов, методу N-грам, пошуку фразеологізмів та аугментації словників через проміжні мови. Також розроблено обробку аудіоданих для визначення мови спікера на основі згорткових нейронних мереж із використанням мелчастотних кепстральних коефіцієнтів (MFCC) та спектрограм.

Третій розділ присвячено методам пошуку слів у словниках, було проведено порівняльний аналіз шести алгоритмів: бінарного пошуку, методу двох вказівників, префіксного дерева, хешування, збалансованого дерева та кореневої декомпозиції. Запропоновано модифікацію префіксного дерева із динамічним масивом замість статичних посилань, що дозволило зняти обмеження по пам'яті та забезпечити лінійне масштабування часу пошуку при збільшенні обсягу словника.

У четвертому розділі розглянуто методи автоматичні метрики оцінки якості перекладу (BLEU, ChrF, METEOR, TER, BLEURT, COMET) та обґрунтовано вибір метрики METEOR як основного критерію оцінювання.

Запропоновані наукові положення та висновки логічно впливають із проведених теоретичних досліджень та підтверджуються результатами моделювання, експериментів і практичної апробації системи у навчальних та виробничих середовищах.

Достовірність та новизна наукових положень, висновків і рекомендацій, сформульованих у дисертації.

Достовірність отриманих результатів забезпечена використанням

сучасних наукових методик та підтверджена експериментальними дослідженнями й моделюванням. До основних наукових результатів, що визначають новизну роботи, належать такі положення:

- вперше запропоновано функціональну модель обробки текстових масивів в гібридній інформаційній системі перекладу штучних мов, яка дозволяє формувати нові словники на основі корпусів текстових даних двома мовами та підвищує точність перекладу штучних мов завдяки інтеграції результатів статистичних і нейромережевих методів обробки голосових та текстових масивів;

- набув подальшого розвитку метод пошуку слів в словниках, в якому, завдяки зміні типу префіксного дерева, забезпечується збільшення продуктивності обчислень при лінійному масштабуванні об'єму текстового словника під час виконання аугментації текстових масивів;

- набув подальшого розвитку метод визначення мови спікера, який забезпечує підвищення точності визначення мови завдяки вибору послідовності методів препроцесінгу вхідних даних, враховуючи різні типи шумів (білий, гаусівський, рожевий, коричневий, імпульсний).

Сформульовані автором висновки та рекомендації коректні, аргументовані та підтверджені результатами моделювання й експериментальних досліджень.

Повнота викладу в наукових публікаціях.

Наукові результати дисертації широко представлені та опубліковані у 15 наукових працях, з них 5 статей у наукових фахових виданнях України категорії А, які індексуються у Scopus, 5 статей у наукових фахових виданнях України категорії Б; 5 тез доповідей у матеріалах міжнародних науково-технічних та науково-практичних конференцій, серед них дві доповіді, що опубліковані в матеріалах міжнародних наукових конференцій, які індексуються у Scopus.

Значущість дослідження для науки і практики.

Запропоновані методи перекладу тексту можуть бути безпосередньо впроваджені в інформаційній системі перекладу штучних мов для покращення якості перекладу тексту та аудіофайлів. Отримані результати є придатними для застосування в онлайн-сервісах дистанційного спілкування та консультування, а також цифрових платформах, де важливою є гарантія достовірності перекладеного тексту у складному акустичному середовищі.

Результати дисертаційної роботи використовуються в освітньому процесі Харківського національного університету радіоелектроніки на кафедрі електронних обчислювальних машин для забезпечення навчальних дисциплін "Комп'ютерні лінгвістичні технології" та "Методи рішення задач загального призначення на графічних процесорах" підготовки здобувачів другого (магістерського) рівня вищої освіти спеціальності F6 Інформаційні системи і технології освітньо-наукової програми Інформаційні системи та технології.

Дисертаційна робота виконана на кафедрі електронних обчислювальних машин Харківського національного університету радіоелектроніки відповідно до плану наукових досліджень у межах госпдоговірної НДР «Підвищення безпеки online консультацій завдяки виявленню фальсифікованого (синтетичного) голосу» (Держреєстраційний номер: 0124U003434), держбюджетної НДР «Мультимодальна система внутрішньої навігації та асистивної підтримки для користувачів з особливими потребами» (Державний реєстраційний номер: 0126U002094) та держбюджетної НТР «Система асистованого супроводу динамічних об'єктів» (Державний реєстраційний номер: 0124U003837) затвердженого Міністерством освіти і науки України. Автор дисертації був одним із виконавців зазначеної науково-дослідної роботи, що свідчить про практичну спрямованість та прикладну цінність отриманих результатів.

Відсутність (наявність) порушення академічної доброчесності.

Під час аналізу дисертації ознак порушення академічної доброчесності виявлено не було. Автор коректно здійснює посилання на використані джерела, результати інших авторів та власні публікації, чітко дотримуючись норм академічної доброчесності.

Дискусійні положення та зауваження щодо змісту дисертаційної роботи і її окремих положень.

1. Обмеженість розглянутих шумів в системі розпізнавання мов ускладнює оцінку практичності запропонованої розподіленої інформаційної системи перекладу, а також вдосконалених та модифікованих методів перекладу тексту.

2. Недостатньо повно розкрито питання масштабованості запропонованої моделі та методів у випадку зростання кількості користувачів і підключення нових підсистем. Зокрема, відсутні експериментальні оцінки продуктивності системи при значному збільшенні навантаження чи кількості вузлів.

3. У роботі бракує систематичного порівняння розробленої технології з існуючими аналогами (як комерційними, так і відкритими рішеннями), що обмежує можливість об'єктивно оцінити переваги запропонованих підходів у практичному застосуванні.

4. У структурі деяких підрозділів дисертаційної роботи наявні повторювані пояснення, що частково ускладнює сприйняття матеріалу. Оптимізація викладення могла б зробити аргументацію більш лаконічною та посилити акцент на ключових технічних досягненнях.

Проте зазначені зауваження не зменшують загальну позитивну оцінку дисертаційного дослідження.

Загальна оцінка роботи, її відповідність встановленим вимогам.

Вважаю, що дисертаційна робота здобувача ступеня доктора філософії

Гаврашенка Антона Олеговича на тему «Модель та методи обробки текстових масивів в інформаційній системі перекладу штучних мов» є закінченим науковим дослідженням, що виконане на належному науковому рівні та не порушує принципів академічної доброчесності. Сукупність отриманих наукових та практичних результатів розв'язує наукове завдання, що має істотне значення для галузі знань Інформаційні технології.

Зважаючи на актуальність, високий рівень наукової новизни, практичну цінність сформульованих положень і результатів, можна зробити висновок, що дисертаційна робота Гаврашенка Антона Олеговича на тему «Модель та методи обробки текстових масивів в інформаційній системі перекладу штучних мов» повністю відповідає пунктам 6, 7, 8, 9 Порядку присудження ступеня доктора філософії (затвердженого Постановою Кабінету Міністрів України від 12 січня 2022 р. №44), а її автор заслуговує на присудження наукового ступеня доктора філософії в галузі знань 12 Інформаційні технології за спеціальністю 126 Інформаційні системи та технології.

Офіційний рецензент:

кандидат технічних наук, доцент

доцент кафедри електронних обчислювальних машин

Харківського національного університету радіоелектроніки

« 29 » 06 2026 р.



Тетяна ФІЛІМОНЧУК

Підпис Філімончук Т.В. засвідчую:

Учений секретар

Харківського національного університету радіоелектроніки

« 29 » 06

2026 р.



 Тетяна ФІЛІМОНЧУК