

РЕФЕРАТ

Розмір пояснювальної записки – 75 аркушів, містить 43 ілюстрації, 25 таблиць, 3 додатки, 38 посилань на джерела.

Актуальність теми. У роботі розглянуто проблематику генерації текстів українською мовою, показано основні особливості наявних інструментів генерування природної мови, їх переваги та недоліки. Виявлено потребу в розробці методу генерації, що інтегрує інструмент морфологічного аналізу, велику мовну модель, модель LDA та інструмент виявлення і виправлення граматичних помилок.

Мета дослідження. Основною метою є підвищення якості згенерованих текстів української мови при заданій тематиці та наборі ключових слів.

Об'єкт дослідження: програмне забезпечення для генерації текстів української мови з використанням великих мовних моделей (LLM) та морфологічних аналізаторів.

Предмет дослідження: методи, програмна архітектура, механізми інтеграції морфологічного аналізу з великими мовними моделями для підвищення якості згенерованих текстів українською мовою.

Для реалізації поставленої мети **сформульовані наступні завдання:**

- аналіз наявних рішень;
- дослідження морфологічних особливостей української мови;
- розробка методу генерації текстів українською мовою;
- реалізувати інтеграцію LLM з морфологічним аналізатором;
- реалізація прототипу програмного забезпечення;
- експериментальне дослідження та аналіз результатів;
- оцінка ефективності запропонованого рішення.

Наукова новизна результатів магістерської дисертації полягає в тому, що було розроблено метод генерації текстів української мови шляхом інтеграції великих мовних моделей з морфологічним аналізатором, застосуванням методу

категоризації LDA та технологією виправлення помилок LanguageTool, що в результаті підвищує якість згенерованих текстів.

Практичне значення отриманих результатів полягає у розробці програмного забезпечення, яке поєднує великі мовні моделі з морфологічними аналізаторами, сприяючи підвищенню якості автоматично згенерованих текстів української мови на задану тематику та набір ключових слів. Це дозволяє розширити можливості застосування такого ПЗ у різноманітних доменах (освіта, наука, медіа, тощо).

Зв'язок з науковими програмами, планами, темами. Робота виконувалась на кафедрі інформатики та програмної інженерії Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського" в рамках теми «Методи та технології високопродуктивних обчислень та обробки надвеликих масивів даних». Державний реєстраційний номер 0117U000924

Апробація. Наукові положення дисертації пройшли апробацію на VI міжнародній науково-практичній конференції молодих вчених та студентів «Інженерія програмного забезпечення і передові інформаційні технології» (SoftTech-2024) – м. Київ.

Публікації. Наукові положення дисертації опубліковані в:

1) Довгополук Р. Р. Метод генерації текстів українською мовою / Р. Р. Довгополук, Ю. О. Олійник. // Матеріали VI Міжнародної науково-практичної конференції молодих вчених та студентів «Інженерія програмного забезпечення і передові інформаційні технології» (SoftTech-2024) – м. Київ: НТУУ «КПІ ім. Ігоря Сікорського», 19-21 травня 2024 р.

Ключові слова: ОБРОБКА ПРИРОДНОЇ МОВИ, УКРАЇНСЬКА МОВА, ГЕНЕРАЦІЯ ТЕКСТУ, ВЕЛИКА МОВНА МОДЕЛЬ, МОРФОЛОГІЧНИЙ АНАЛІЗ, ТЕМАТИЧНЕ МОДЕЛЮВАННЯ.