

УДК 004

Белименко Д.Р.¹, Олійник А.О.²

¹ студ. гр. КНТ-114м НУ «Запорізька політехніка»

² проф. НУ «Запорізька політехніка»

ДОСЛІДЖЕННЯ ТА ПРОГРАМНА РЕАЛІЗАЦІЯ ЗАСОБІВ РОЗПІЗНАВАННЯ ТА ПЕРЕКЛАДУ АУДІОПОТОКУ В РЕАЛЬНОМУ ЧАСІ

Об'єкт дослідження – процес розробки програмного забезпечення для синхронного перекладу аудіопотоку в реальному часі.

Предмет дослідження – методи та програмні засоби інтеграції нейромережових моделей для локальної обробки мовлення з низькою затримкою.

Мета роботи – розробка програмного забезпечення для захоплення, транскрибації та перекладу системного аудіо в режимі реального часу, що дозволяє зменшити час розуміння іншомовного аудіоконтенту користувачем.

Технічні засоби: мова програмування Python, бібліотеки Faster-Whisper, Transformers (MarianMT), CTranslate2, PyTorch, SoundDevice, BlackHole, бібліотека графічного інтерфейсу Tkinter.

У результаті виконання роботи було створено повністю автономну систему синхронного перекладу аудіопотоку.

На етапі аналізу існуючих рішень було встановлено, що більшість комерційних та відкритих систем, призначених для перекладу в реальному часі, мають критичні недоліки, пов'язані з необхідністю постійного використання хмарних сервісів та API. Це створює прямі ризики щодо конфіденційності оброблюваних голосових даних, а також робить систему вразливою до перебоїв в інтернет-з'єднанні.

Розроблений програмний комплекс Real-Time Audio Translator успішно вирішує ці проблеми, реалізуючи принцип стопроцентної локальної обробки даних. Уся архітектура побудована на оптимальних технологіях, які поєднують захоплення системного аудіо через BlackHole, розпізнавання мовлення моделлю Whisper та нейромережвий переклад за допомогою MarianMT, при цьому всі операції виконуються безпосередньо на пристрої користувача.

Наукова новизна одержаних результатів:

- удосконалено архітектуру конвеєра обробки аудіо шляхом інтеграції віртуального аудіокабелю (BlackHole) з алгоритмом виявлення мовленнєвої активності (VAD), що базується на комбінованому аналізі амплітуди та частотного спектру, що дозволяє зменшити обчислювальне навантаження на 40-60%;

- запропоновано дворівневу систему кешування перекладів, яка поєднує словниковий пошук та LRU-кеш для нейромережових моделей (MarianMT), що забезпечує скорочення часу повторного перекладу з 1.2 с до 0.1 с;

- удосконалено підхід до керування ресурсами під час роботи з великими мовними моделями шляхом реалізації динамічного вивантаження моделей (LRU eviction) та автоматичної адаптації під обчислювальний пристрій (Apple Metal / NVIDIA CUDA / CPU), що дозволяє використовувати систему на персональних комп'ютерах з обмеженим обсягом пам'яті;

- розроблено інтегроване програмне середовище, що об'єднує процеси транскрибації (Whisper) та перекладу (OPUS-MT) в єдиному локальному циклі без передачі даних третім сторонам.

Наукова новизна підтверджується не лише інтеграцією цих сучасних моделей, а й удосконаленням алгоритмів оптимізації. Зокрема, було розроблено багатофакторний детектор голосової активності VAD, що мінімізує обчислювальне навантаження, ігноруючи тишу та сторонні шуми. Ключовим досягненням є також впровадження дворівневої системи кешування з використанням стратегії LRU, що дозволило скоротити час перекладу повторюваних фраз до 0.09 – 0.2 с, забезпечуючи необхідну швидкість для роботи в режимі реального часу.

Виконано проєктування та реалізація програмного комплексу Real-Time Audio Translator, проведено тестування продуктивності на різних апаратних платформах, підтверджено ефективність локальної обробки даних.

Проведена експлуатаційна перевірка та тестування підтвердили, що програмне забезпечення є функціонально коректним, стабільним і демонструє високу продуктивність завдяки використанню апаратного прискорення Metal/CUDA.

Галузь використання – програмний застосунок може бути застосований для споживання іншомовного медіа-контенту, проведення міжнародних онлайн-конференцій, а також в інклюзивній освіті для автоматичного субтитрування лекцій.