

РОЗРОБКА ПРОГРАМНОЇ СИСТЕМИ ДЛЯ АНАЛІЗУ ГОЛОСОВИХ ПОВІДОМЛЕНЬ

¹Вінницький національний технічний університет

Анотація

Запропоновано підхід до розробки desktop-застосунку для аналізу голосових повідомлень із використанням засобів штучного інтелекту, який поєднує автоматичну транскрибацію мовлення на основі моделі Whisper з інтелектуальним аналізом тексту через генеративну модель ChatGPT та формуванням структурованого аналітичного звіту, що дозволяє підвищити ефективність опрацювання аудіоконтенту.

Ключові слова: аналіз голосових повідомлень, штучний інтелект, транскрибація мовлення, Whisper, ChatGPT, desktop-застосунок.

Abstract

An approach to the development of a desktop application for voice message analysis using artificial intelligence tools is proposed, which combines automatic speech transcription based on the Whisper model with intelligent text analysis via the generative ChatGPT model and the generation of a structured analytical report, thus improving the efficiency of audio content processing.

Keywords: voice message analysis, artificial intelligence, speech transcription, Whisper, ChatGPT, desktop application.

Вступ

Сучасний розвиток цифрових технологій та засобів штучного інтелекту (ШІ) створює нові можливості для автоматизації обробки інформації у різних сферах діяльності. Одним з актуальних напрямів є аналіз голосових повідомлень, що дозволяє перетворювати аудіоконтент у структуровані текстові звіти, виділяти ключові теми, тези та цитати. Розвиток технологій автоматичного розпізнавання мовлення (ASR) та генеративного ШІ відкриває можливості для побудови комплексних інструментів, які не лише розпізнають мовлення, а й виконують його смисловий аналіз [1].

Метою роботи є розроблення desktop-застосунку для аналізу голосових повідомлень із інтегрованими засобами штучного інтелекту, що поєднує автоматичну транскрибацію мовлення з інтелектуальним аналізом тексту для підвищення ефективності опрацювання аудіоінформації.

Результати дослідження

Аналіз існуючих програмних рішень (Otter.ai, Fireflies.ai, Whisper, Google Speech-to-Text, Notta.ai) показав, що більшість з них є хмарними сервісами без десктопної версії, мають обмежену підтримку української мови та виконують переважно транскрибацію без подальшого смислового аналізу. Жоден з проаналізованих аналогів не поєднує транскрибацію з формуванням структурованого аналітичного звіту, що підтверджує актуальність розробки [2].

Запропонований підхід базується на монолітній багаторівневій архітектурі з інтеграцією хмарних API. Графічний інтерфейс реалізовано мовою Java з використанням бібліотеки JavaFX, а серверну логіку – мовою Python. Взаємодію між Java-інтерфейсом та Python-бекендом забезпечує клас PythonBridge на основі механізму ProcessBuilder. Для транскрибації використано API OpenAI Whisper, для смислового аналізу – API ChatGPT, а для збереження результатів – СУБД SQLite.

Структуру програмної системи представлено у вигляді п'яти функціональних модулів: модуль запису та завантаження аудіо, модуль транскрибації, модуль інтелектуального аналізу, модуль графічного інтерфейсу та модуль збереження даних. Узагальнена схема взаємодії модулів зображена на рис. 1.

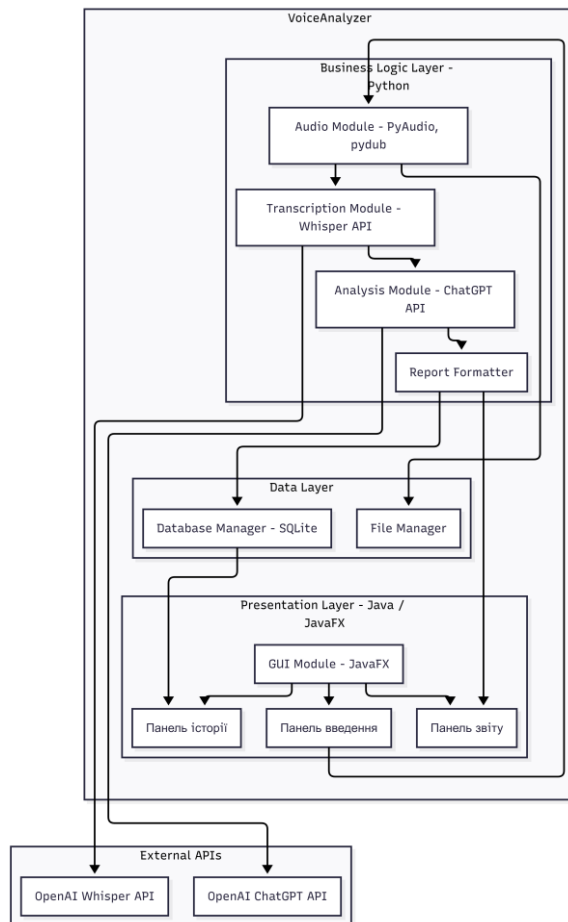


Рис. 1. Структура програмної системи аналізу голосових повідомлень

Обробка голосового повідомлення реалізована у вигляді конвеєра: після запису з мікрофона або завантаження аудіофайлу (формати WAV, MP3, OGG) виконується валідація розміру з урахуванням обмеження API Whisper у 25 МБ та за потреби автоматичне розбиття на сегменти. Транскрибований текст передається до API ChatGPT з промптом, що визначає структуру звіту: основна тема, ключові тези, занотовані деталі, цитати та висновок.

Точність розпізнавання P визначається як відношення кількості правильно розпізнаних слів до їх загальної кількості:

$$P = N_{\text{правильних}} / N_{\text{загальних}} \cdot 100\%, \quad (1)$$

де $N_{\text{правильних}}$ – кількість правильно розпізнаних слів;

$N_{\text{загальних}}$ – загальна кількість слів у тестовому записі.

Експериментальні дослідження підтвердили, що використання моделі Whisper забезпечує коректну транскрибацію україномовного та англійськомовного контенту з автоматичним визначенням мови, а інтеграція з ChatGPT дозволяє формувати структуровані звіти, обсяг яких автоматично масштабується залежно від тривалості вхідного запису.

Висновки

Запропоновано підхід до розробки desktop-застосунку для аналізу голосових повідомлень із використанням засобів штучного інтелекту, що дозволяє об'єднати автоматичну транскрибацію мовлення з інтелектуальним аналізом тексту. Розроблена програмна система забезпечує формування

структурованого аналітичного звіту з визначенням основної теми, ключових тез, деталей та цитат, підтримує україномовний та англійськомовний контент і зберігає історію аналізів у локальній базі даних, що підвищує ефективність опрацювання аудіоінформації у бізнесі, освіті та повсякденній діяльності.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Radford A. Robust Speech Recognition via Large-Scale Weak Supervision / A. Radford, J. W. Kim, T. Xu et al. // arXiv preprint arXiv:2212.04356. – 2022.
2. OpenAI API [Електронний ресурс]. – Режим доступу: <https://platform.openai.com/docs> (дата звернення: 25.05.2026).
3. JavaFX Documentation [Електронний ресурс]. – Режим доступу: <https://openjfx.io> (дата звернення: 25.05.2026).

Владислав Олександрович Антонюк – студент групи ЗПІ-226, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: antonukvlad70@gmail.com

Науковий керівник: **Рейда Олександр Миколайович** – канд. техн. наук, доцент кафедри програмного забезпечення, Вінницький національний технічний університет, м. Вінниця

Vladyslav O. Antoniuk – Department of Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: antonukvlad70@gmail.com.

Supervisor: **Oleksandr M. Reida** – Cand. Sc. (Eng), Assistant Professor of Software Engineering Department, Vinnytsia National Technical University, Vinnytsia