

ІНФОРМАЦІЙНА СИСТЕМА ДЛЯ РОЗПІЗНАВАННЯ МОВЛЕННЯ У ЗВУКОВИХ ФАЙЛАХ

Драч А.О, науковий керівник - к.ф.-м.н., доцент Кириченко В.В.

У сучасному інформаційному суспільстві обсяг аудіоінформації стрімко зростає. Звукові файли, які містять мовлення – це записи дзвінків, інтерв'ю, конференцій, лекцій, голосові повідомлення тощо – активно використовуються в різних сферах: журналістиці, освіті, медицині, юриспруденції, бізнесі та державному управлінні. Проте для ефективної роботи з таким контентом потрібні зручні засоби автоматичного розпізнавання мовлення та його перетворення у текстовий формат.

Інформаційні системи для розпізнавання мовлення дозволяють значно знизити витрати часу на транскрибування аудіозаписів, забезпечити швидкий доступ до змісту усних повідомлень, а також відкривають нові можливості для аналізу даних, пошуку інформації та автоматизації рутинних завдань. Особливо актуальною ця тема є у зв'язку з розвитком штучного інтелекту, хмарних сервісів та машинного навчання, які дозволяють створювати все більш точні й універсальні системи розпізнавання мовлення.

Зважаючи на зростаючі вимоги до швидкості та якості обробки звукових даних, створення та вдосконалення інформаційних систем для розпізнавання мовлення набуває все більшого значення. Це є важливим кроком у напрямку цифрової трансформації, підвищення ефективності роботи з великими обсягами інформації та розширення можливостей взаємодії людини з комп'ютером.

Навчання моделей для розпізнавання мовлення вимагає великих масивів аудіоданих з точними текстовими транскрипціями. Окрім того, важливо використовувати адаптивні мовні моделі, які можуть покращити якість розпізнавання шляхом врахування граматичного та семантичного контексту.

У роботі розглядається програмна модель, яка здатна ефективно трансформувати дані із аудіоформату у текст. Для реалізації такої системи потрібно створити.

1.Клієнтська частина

Це фронтенд-компонент, адаптований для роботи на мобільних пристроях (Android/iOS).

2. Бекенд (серверна частина)

Центральний елемент системи, який забезпечує обробку аудіоданих і управління логікою застосування.

3. Графічний веб-інтерфейс

Інтуїтивно зрозумілий вебінтерфейс, що дозволяє користувачам:

Для досягнення високої точності система повинна мати щільну інтеграцію між акустичною моделлю (яка розпізнає звуки) та мовною моделлю (яка формує правильний текст на основі контексту).

Глибокі нейронні мережі, зокрема рекурентні та трансформерні архітектури, можуть бути використані для обробки послідовностей звукових даних та побудови найбільш ймовірного текстового представлення. У системі також передбачається використання **нейронної мережі**, призначеної для **подавлення шумів**, які можуть негативно впливати на точність перетворення аудіосигналів у текст. Ця модель буде виконувати функцію **попередньої обробки (преобробки)**, фільтруючи фонові завади та покращуючи якість мовного сигналу перед його подачею на основний модуль розпізнавання.

Процес роботи інтелектуальної системи для розпізнавання мовлення включає такі етапи:

- 1. Підготовка даних:** збирання великого набору аудіофайлів з точними текстовими транскрипціями.

2. **Створення нейронної мережі:** побудова моделі, здатної розпізнавати мовні патерни, акценти, шум та інтонацію.
3. **Навчання моделі:** оптимізація параметрів на основі навчального набору, з урахуванням як акустичних, так і мовних аспектів.
4. **Тестування та оцінка:** перевірка системи на нових аудіофайлах для визначення точності розпізнавання та виявлення потенційних помилок.

Таким чином, інтелектуальна система розпізнавання мовлення у звукових файлах дозволить автоматизувати обробку аудіоінформації, забезпечити швидкий і точний перехід від усного мовлення до тексту, та має потенціал для застосування у широкому спектрі практичних задач – від освіти до бізнесу та медицини.

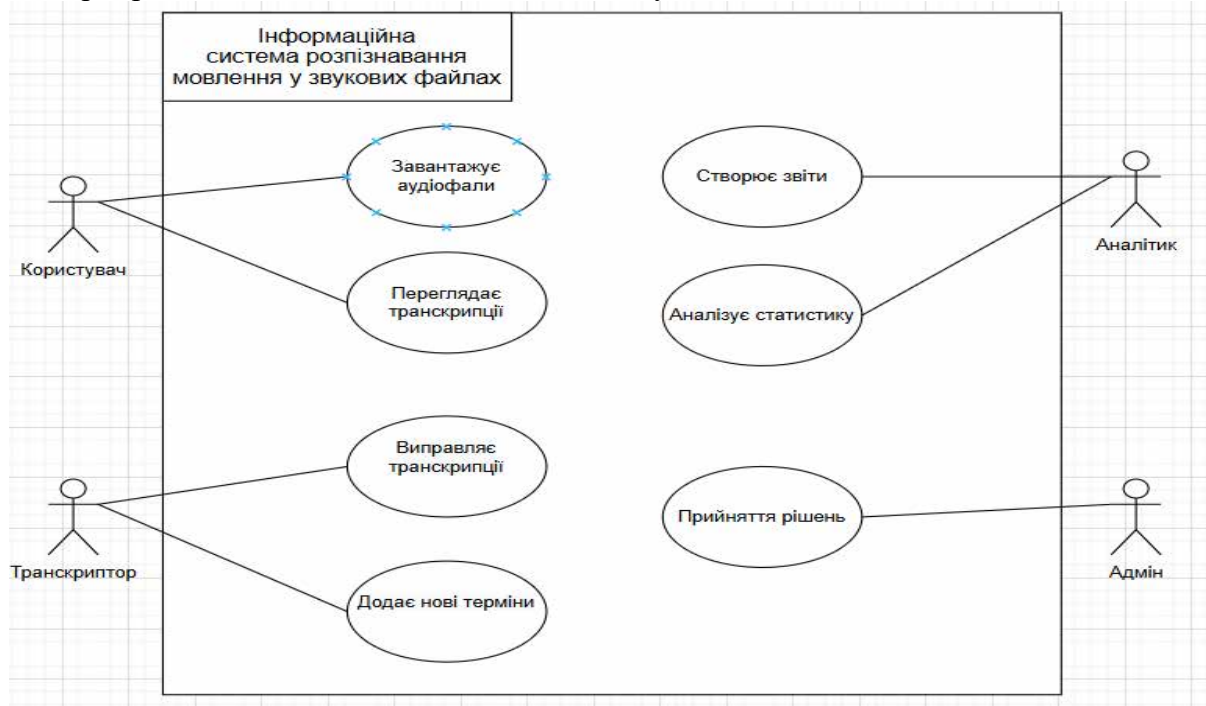


Рис-1. Діаграму прецедентів

У результаті реалізації інтелектуальної системи розпізнавання мовлення, що функціонує за запропонованою архітектурою, аудіозаписи успішно трансформуються у текстовий формат.

Це стало можливим завдяки використанню нейронних мереж, здатних адаптуватися до особливостей голосу, темпу мовлення, акцентів та рівня шуму. Поєднання акустичних моделей з мовними моделями дозволяє досягти високої точності розпізнавання мовлення. При подальшій інтеграції із системами автоматичного перекладу можливо отримати якісний переклад розпізнаного мовлення іншими мовами.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Mozilla DeepSpeech Project (<https://github.com/mozilla/DeepSpeech>) — Відкрите ПЗ для розпізнавання мовлення.
2. OpenAI Whisper (<https://github.com/openai/whisper>) — сучасна модель розпізнавання мовлення з відкритим кодом.
3. Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing* (3rd Edition, Draft).

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**



**ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ
АСПЕКТИ РОЗРОБКИ
КОМП'ЮТЕРНИХ СИСТЕМ**

**ЗБІРНИК НАУКОВИХ ПРАЦЬ ЗА МАТЕРІАЛАМИ
VII ВСЕУКРАЇНСЬКОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ
СТУДЕНТІВ І АСПІРАНТІВ
*24 квітня 2025 року***

Київ 2025

УДК 004

Відповідальна за випуск: М.І. Лендел

Збірник наукових праць за матеріалами VII Всеукраїнської науково-практичної конференції студентів і аспірантів «ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ АСПЕКТИ РОЗРОБКИ КОМП'ЮТЕРНИХ СИСТЕМ '2025», 24 квітня 2025 року, НУБіП України, Київ. – 452 с. (електронне видання)

Відповідальність за зміст публікацій несуть автори.

Передрук матеріалів, а також використання їх будь-якій формі допускається лише з дозволу авторів