

Сергій Сергієнко, Проць Андрій*

¹ НЛТУ України, Львів, Україна, 122.24serhiienko.s@nltu.lviv.ua

² НЛТУ України, Львів, Україна, andrii_prots@nltu.edu.ua

Інформаційна система для транскрибування аудіозаписів

***Анотація.** Інформаційна система призначена для автоматичного перетворення мовлення на текст. Вона реалізована мовою програмування Python з використанням бібліотек Tkinter для графічного інтерфейсу та SpeechRecognition для розпізнавання. Ключовим функціоналом є можливість транскрибування з локальних аудіо/відео файлів, живого запису з мікрофона та аудіодоріжок YouTube відео. Розроблена система є ефективним інструментом, що підвищує продуктивність роботи з мультимедійним контентом, мінімізуючи ручне введення даних.*

***Ключові слова** – транскрибування мовлення, розпізнавання мови, Python, Google Speech-to-Text.*

Вступ. Розробка систем для автоматичного розпізнавання аудіозаписів є важливою через зростаючий об'єм голосової інформації у різних сферах. Актуальність проекту обумовлена потребою автоматизації процесу аналізу мовних даних, що підвищує ефективність обробки інформації. Технології розпізнавання мови Google Speech-to-Text забезпечують високу якість транскрибування аудіо у текст. Розробка зручного графічного інтерфейсу спрощує взаємодію користувача з системою і розширює коло потенційних користувачів. Система може бути застосована у багатьох галузях, таких як журналістика, освіта, медицина, бізнес, де важливо працювати з аудіоконтентом. Застосування автоматизованих інструментів для обробки аудіо зменшує затрати часу і людських ресурсів на цей процес.

Розробка інтелектуальної системи. Мета роботи – розробка та програмна реалізація інформаційної системи для транскрибування відео та аудіозаписів. Сутність цього як завдання це інтерпретація акустичного сигналу, виявлення мовних одиниць, їх зіставлення з лексичними одиницями мови та побудова по-

слідовності слів у вигляді тексту. Це передбачає проходження кількох ключових етапів [1].



Рисунок 1. Ключові етапи автоматичного транскрибування аудіозаписів

У процесі реалізації інформаційної системи ключову роль відіграють джерела вхідної інформації, які визначають як якість кінцевого результату, так і загальну ефективність системи. Вхідна інформація – це сукупність даних, що надходять на вхід програмного забезпечення та обробляються для подальшої генерації текстової інформації.

Головним джерелом вхідної інформації є аудіозаписи, які користувачі завантажують до системи. Це можуть бути записи лекцій, інтерв'ю, конференцій, телефонних розмов, відеозаписів із розмовною складовою. Формати аудіо можуть бути різними: .mp3, .wav. Якість аудіофайлу, тривалість, наявність шумів та акценти мовців напряду впливають на результат розпізнавання [2].

У разі створення моделі розпізнавання мови система потребує навчальних наборів даних, розмічених аудіозаписів із відповідними текстовими розпізнаваннями. Прикладами таких джерел є відкриті набори даних LibriSpeech, Common Voice, TED-LIUM, UkrVoice для української мови. Ці ресурси дозволяють навчити систему розпізнавати мовлення певними мовами з різними акцентами та у різних акустичних умовах.

Вхідними даними системи є аудіофайли або аудіопотоки, що містять мовлення. Дані можуть бути надані користувачем або отримані з зовнішніх сервісів. Система може приймати метадані: назву файлу, дату та час запису, ідентифікатор користувача, опис або мітки запису. Ці метадані дозволяють класифікувати та впорядковувати транскрибовані записи. Чітке структурування

вхідних і вихідних даних є фактором успішного функціонування системи, забезпечення її точності, масштабованості та зручності для користувача.

У процесі перетворення голосу на текст система виявляє як окремі слова, так і цілі фрази в аудіовході, що надходить від людини чи машини, а потім перетворює їх на читабельний текстовий формат. Ця можливість забезпечує ефективну комунікацію в трьох основних напрямках: між людиною та людиною, людиною та машиною, а також машиною та людиною. Особливу цінність така технологія має в ситуаціях, коли спілкуються або взаємодіють між собою особи, які користуються різними мовами чи діалектами. Без системи перетворення вони можуть просто не розуміти слів, сказаних одне одним. Відповідно у таких випадках конвертер стає помічником, оскільки він розпізнає слова, вимовлені людиною з незнайомим акцентом чи діалектом і трансформує їх у текстову форму, яку іншій особі легко прочитати та зрозуміти [3].

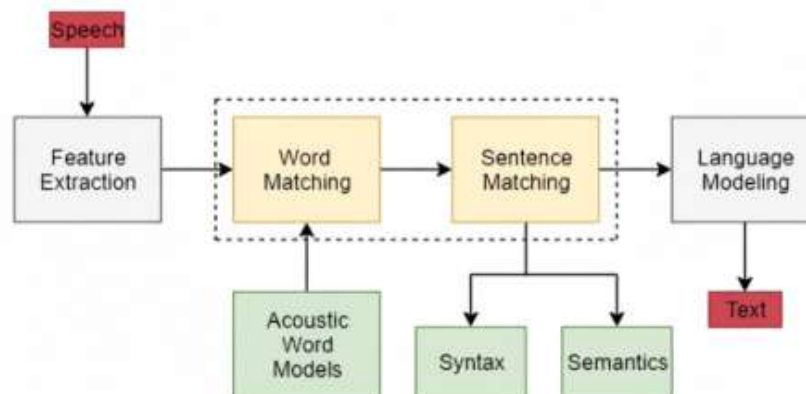


Рисунок 2. Процес перетворення мовлення на текст

Задача розпізнавання аудіоінформації є складною, багаторівневою та математично обґрунтованою задачею, що вимагає залучення методів обробки сигналів, машинного навчання та лінгвістичного аналізу для досягнення високої точності. Математичне моделювання у задачі транскрибування аудіоінформації базується на комбінації ймовірнісних моделей, методів статистичної обробки сигналу, глибокого навчання та лінгвістичного аналізу. Ефективне використання цих методів дозволяє досягти високої точності перетворення мовлення в текст.



Рисунок 3. Етапи роботи з даними

Побудова та вибір ознак є ключовим етапом у процесі розв’язання задачі транскрибування аудіоінформації, оскільки саме на основі якісних ознак відбувається навчання моделей машинного або глибокого навчання. Ознаки служать представленням вхідних аудіосигналів, що дозволяє моделі виділяти та узагальнювати інформацію, яка необхідна для точного перетворення мовлення в текст.

Для розробки системи транскрибування аудіоінформації було обрано мову програмування Python. Основні причини вибору полягають у підтримці бібліотек машинного навчання, простоті синтаксису, наявності ефективних інструментів для роботи з аудіо та мовними моделями.



Рисунок 4. Бібліотеки машинного навчання Python

Для того, щоб користувач міг взаємодіяти з інформаційною системою, в роботі використано бібліотеку Tkinter. Вона відповідає за створення графічного інтерфейсу (GUI) інформаційної системи, надаючи кнопки для завантаження файлів, відображення тексту та управління процесом транскрибування.

Експериментальна частина. Інтерфейс інтелектуальної системи представлено нижче.

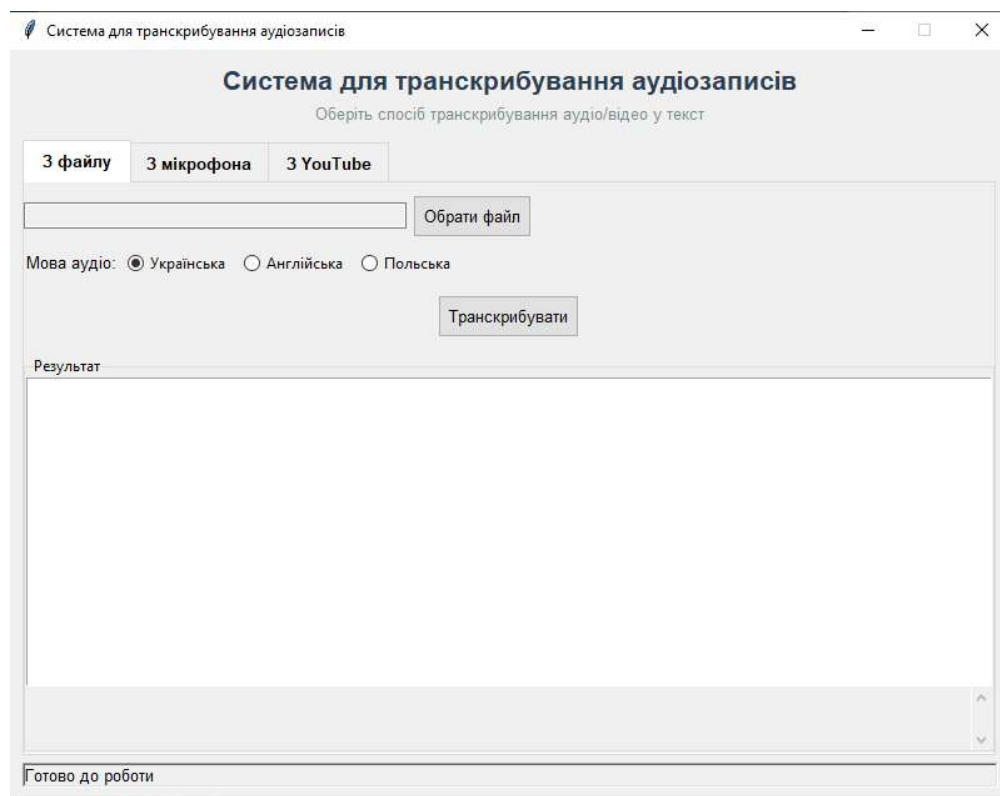


Рисунок 5. Інтерфейс інформаційної системи для транскрибування аудіозаписів

Розроблена інформаційна система забезпечує користувачу простий та інтуїтивно зрозумілий інтерфейс для розпізнавання мовлення з трьох основних джерел: локального файлу, мікрофона та зовнішнього YouTube відео. Робота користувача з системою організована через три вкладки, що відповідають трьом режимам транскрибування.

Режим транскрибування з локального файлу призначений для обробки попередньо записаних аудіо та відеофайлів, що зберігаються на комп'ютері користувача. Цей режим з мікрофона реалізує функцію розпізнавання мовлення в реальному часі для створення нових текстових документів. Він дозволяє з YouTube розпізнавати аудіодоріжку з відео, доступного в мережі інтернет.

Висновок. Розроблено багатофункціональну інформаційну систему для транскрибування аудіозаписів, яка здатна обробляти дані з різних джерел. Було проведено аналіз сучасних методів розпізнавання мовлення та обрано підхід, що базується на використанні хмарних API та локальних інструментів. Для реалізації було обрано мову програмування Python через її бібліотеки у сфері обробки даних та штучного інтелекту. Графічний інтерфейс користувача розроблено за допомогою бібліотеки Tkinter, забезпечуючи кросплатформну роботу та простоту використання. Система успішно інтегрує три ключові режими роботи: транскрибування з локальних файлів, запис і розпізнавання з мікрофона та обробка YouTube-відео. Використання сервісу Google Speech Recognition API забезпечує високу точність для підтримуваних мов (української, англійської, польської).

Список використаних літературних джерел

- [1] Jelinek, F. Statistical Methods for Speech Recognition. MIT Press, 1997. 283 p.
- [2] Rabiner, L. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE. 1989. Vol. 77(2). p. 257-286.
- [3] Graves, A., Mohamed, A., & Hinton G. Speech recognition with deep recurrent neural networks. ICASSP, 2013. p. 6645-6649.

Information system for transcribing audio recordings

Serhiienko Serhii, Prots Andrii

The information system is designed for automatic speech-to-text conversion. It is implemented in the Python programming language using the Tkinter libraries for the graphical interface and SpeechRecognition for recognition. The key functionality is the ability to transcribe from local audio/video files, live recording from a microphone, and audio tracks of YouTube videos. The developed system is an effective tool that increases the productivity of working with multimedia content, minimizing manual data entry.