

СЕКЦІЯ 4. ПРОГРАМНА ІНЖЕНЕРІЯ ТА МОДЕЛЮВАННЯ СКЛАДНИХ РОЗПОДІЛЕНИХ СИСТЕМ

УДК 004.41

А. Бачинський; А. Бачинський

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

ДОСЛІДЖЕННЯ ЗАСОБІВ РОЗРОБКИ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ КОНВЕРТАЦІЇ МЕДІА ФАЙЛІВ У ТЕКСТОВИЙ ФОРМАТ

UDC 004.41

A. Bachynskiy; A. Bachynskiy

RESEARCH OF SOFTWARE DEVELOPMENT TOOLS FOR CONVERTING MEDIA FILES INTO TEXT FORMAT

Сучасне суспільство стикається з постійно зростаючим обсягом мультимедійних даних, які потребують ефективної обробки та збереження. Конвертація аудіо та відео в текстовий формат має велике значення для підвищення доступності інформації, створення баз даних, аналізу тексту та автоматизації рутинних процесів. Важливість цього напрямку підтверджується широким використанням у журналістиці, освіті, судовій практиці, медицині та інших сферах [1].

В даній доповіді пропонується фреймворк для дослідження проєктів з розробки програмних продуктів з метою оцінювання ефективності процесів розробки з впровадженням CI/CD. Для досягнення цієї мети пропонується оцінювати процеси за такими характеристиками [2].

1. Проаналізувати сучасні алгоритми розпізнавання мовлення.
2. Розробити програмне забезпечення, яке інтегрує найкращі методи розпізнавання.
3. Забезпечити підтримку різних форматів файлів(mp3,wav,mp4 тощо).
4. Тестувати систему на різних типах даних для оцінки ефективності.

У роботі реалізовано систему конвертації медіа-файлів та аудіозаписів у текстовий формат із застосуванням сучасних веб-технологій. Використання JavaScript як основної мови розробки забезпечує високу продуктивність та гнучкість коду. Фреймворк Vue дозволив створити інтерактивний інтерфейс для взаємодії з користувачем, зокрема завантаження файлів, налаштування параметрів обробки та виводу результатів. Основна перевага такого підходу – кросплатформність і можливість використання системи у веб-браузері без додаткових програм [3].

Для обробки аудіо- та відеофайлів інтегровано сторонні API та бібліотеки розпізнавання мовлення, які підтримують багатомовність та обробку у реальному часі. У роботі використано підхід клієнт-серверної архітектури: клієнтська частина на Vue відповідає за взаємодію з користувачем, а серверна обробляє файли за допомогою API, таких як Google Speech-to-Text чи Whisper[4]. Це дозволило мінімізувати затримки та досягти високої точності розпізнавання (WER – до 10% для чистого аудіо). Стек технологій для розробки наведено у таблиці 1.

Розроблене програмне забезпечення може використовуватися як основа для створення комерційних сервісів, покращення доступності інформації для людей з інвалідністю, автоматизації аналітики в бізнесі[5]. Окрім користі для людей з інвалідністю та автоматизації бізнес-процесів, цей проєкт може бути корисним у таких сферах:

1. Освіта. Автоматична трансформація записів лекцій у текстовий формат допомагає студентам створювати конспекти, особливо коли немає можливості записувати вручну.
2. Журналістика та медіа. Журналісти можуть використовувати ваш інструмент для автоматизації процесу транскрипції аудіо- та відеоінтерв'ю, що значно економить час. Розпізнавання тексту у відео та подкастах допомагає створювати архіви для подальшого пошуку та аналізу [6].

Таблиця 1. Огляд ключових технологій

Компонент	Технологія	Функція	Переваги
Фронтенд	JavaScript, Vue.js	Створення інтерактивного користувацького інтерфейсу	Кросплатформність, швидка розробка
Серверна частина	SQL	Обробка запитів, управління медіа-файлами	Легкість інтеграції з API
Розпізнавання мовлення	Google Speech-to-Text API	Конвертація аудіо у текст	Висока точність, багатомовність
Обробка файлів	FFmpeg	Конвертація форматів файлів (mp3, wav, mp4)	Підтримка різноманітних форматів

3. Судова система. Автоматичне створення протоколів із записів судових засідань зменшує навантаження на працівників канцелярій та покращує точність записів.

4. Здоров'я та медицина. Записи лікарів, зроблені у вигляді аудіо, можуть автоматично перетворюватися в текст для використання в електронних медичних картках. Інструмент дозволяє перетворювати аудіозаписи консультацій у текстовий вигляд, спрощуючи створення звітів.

5. Розробка програмного забезпечення. Інструмент може бути корисним для ІТ-команд для створення протоколів із записів онлайн-зустрічей, наприклад, у Zoom чи Google Meet.

6. Особисте використання. Люди можуть використовувати систему для конвертації голосових нотаток у текст, полегшуючи особисту організацію. Транскрипція допомагає блогерам і подкастерам створювати текстові версії своїх матеріалів, що позитивно впливає на SEO.

Розробка систем для конвертації аудіо та відео в текстовий формат є важливим напрямом в сучасній інформатиці. Створений прототип підтверджує доцільність використання новітніх алгоритмів розпізнавання мовлення для забезпечення високої точності результатів.

Література

1. Digital Media Literacy Education Framework - An overview of concepts and techniques involving digital media management and file handling (ERIC, 2022)
2. Modern Patterns for Developing Modern Applications by Alex Banks and Eve Porcello – A book to dive into React.js.
4. "JavaScript: The Definitive Guide" by David Flanagan – A comprehensive guide to JS, including modern features(2023-2024).
5. RESTful API Design with Node.js 10 practical book for creating APIs using Node.js. "Designing Web APIs: Building APIs That Developers Love" by Brenda Jin, Saurabh Sahni, and Amir Shevat - API Design Guide.
6. "Technology and Its Impact on Modern Communication: Case Studies on Video Conferencing" – Analysis of video platforms and their impact on society (Cambridge Core, 2023)
7. "Speech and Language Processing" by Daniel Jurafsky and James H. Martin – A primer on speech recognition and natural language processing.
1. Deep Learning for Speech Recognition Review of Machine Learning Algorithms Applied to Speech (Springer).