TRƯỜNG ĐẠI HỌC ĐẠI NAM
KHOA CÔNG NGHỆ THÔNG TIN

Hệ Thống Nhận Diện Giọng Nói & Xác Định Người Nói
(Real-Time Voice Recognition & Speaker Identification System)

Lê Xuân Dương, Nguyễn Công Uẩn, Bùi Anh Tuấn, Đỗ Huy Dũng

Giảng viên hướng dẫn: Lê Trung Hiếu, Nguyễn Thái Khánh

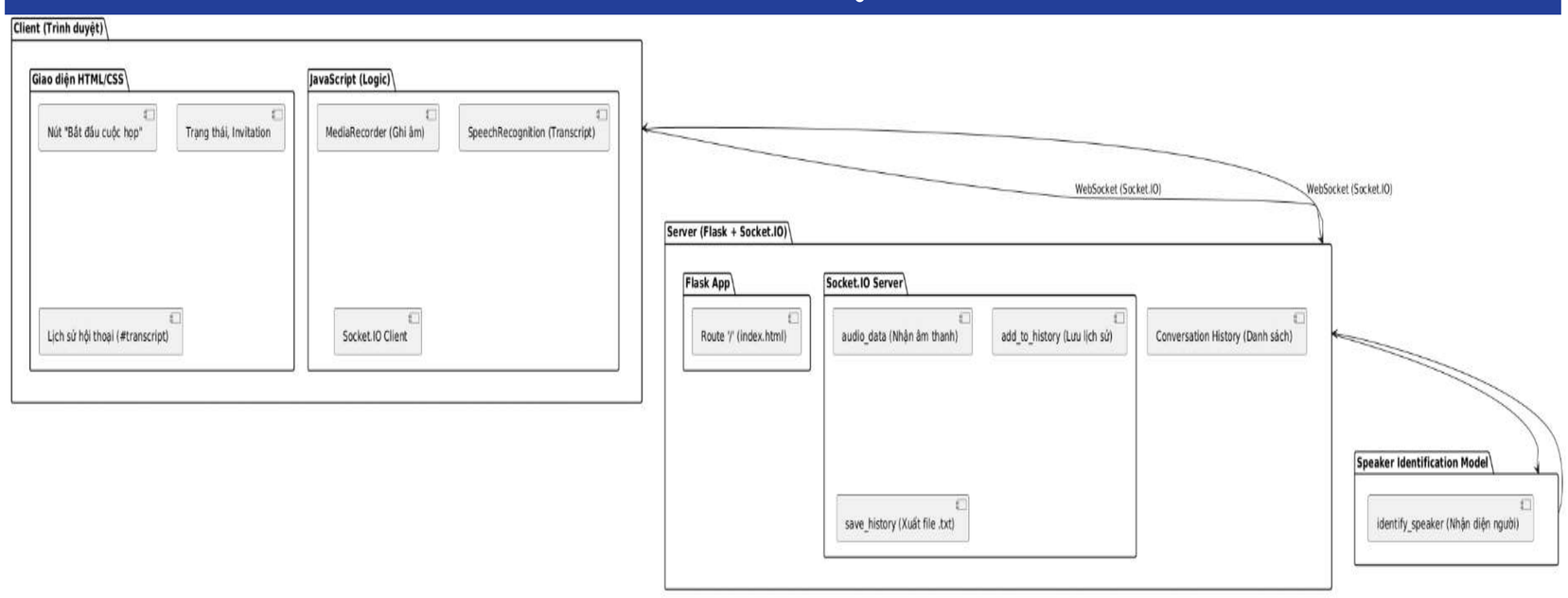
Github: https://github.com/CongUan04/Nhan_Dien_Giong_Noi.git

GIỚI THIỆU

Bối cảnh: Trong thời đại hội họp trực tuyến và giao tiếp từ xa ngày càng phổ biến, việc tự động ghi nhận nội dung cuộc họp và phân biệt người nói đóng vai trò quan trọng trong quản lý thông tin và hỗ trợ các ứng dụng thông minh.

Mục tiêu: Xây dựng một hệ thống nhận diện giọng nói theo thời gian thực kết hợp với việc xác định người nói, qua đó lưu trữ transcript của các cuộc họp và tạo điều kiện cho các ứng dụng xử lý dữ liệu âm thanh.

KIẾN TRÚC HỆ THỐNG



Phương pháp sử dụng

Quy trình xử lý

1. Khởi động

Server: Flask và SocketIO khởi tạo, lưu lịch sử trong conversation_history.

Client: Tải giao diện, kết nối SocketIO, hiển thị nút "Bắt đầu cuộc họp".

2. Ghi âm và nhận diện

Nhấn "Bắt đầu cuộc họp":

Truy cập micro, khởi tạo MediaRecorder và SpeechRecognition.

Gửi audio định kỳ (5 giây) qua socket.emit('audio_data').

Server: Nhận audio, gọi identify_speaker, gửi speaker_result về client.

Client: Gán identifiedSpeaker, hiển thị lời mời phát biểu.

3. Xử lý transcript

Nhận diện giọng nói:

SpeechRecognition lưu transcript vào currentTranscript.

Kiểm tra từ khóa:

"Kết thúc":

Loại "kết thúc" khỏi currentTranscript.

Ghi Tên người nói: Nội dung vào #transcript.

Gửi add_to_history tới server.

Reset, hiển thị "Mời người tiếp theo phát biểu", tiếp tục nhận diện.

"Kết thúc cuộc họp":

Loại "kết thúc cuộc họp" khỏi currentTranscript.

Ghi vào #transcript.

Gửi add_to_history và save_history tới server.

Dừng ứng dụng.

4. Xuất lịch sử

Server: Nhận save_history, tạo file .txt, gửi download_history.

Client: Tạo link tải file tự động.

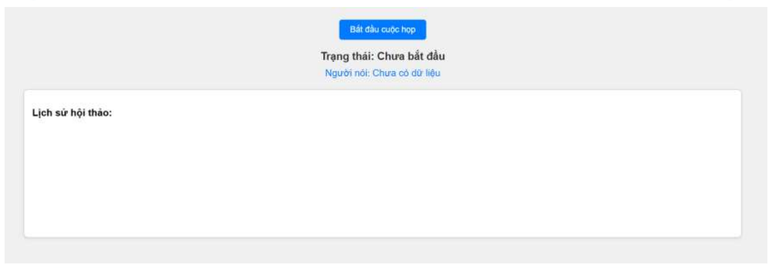
5. Dừng

Dừng MediaRecorder, SpeechRecognition, cập nhật trạng thái "Kết thúc".

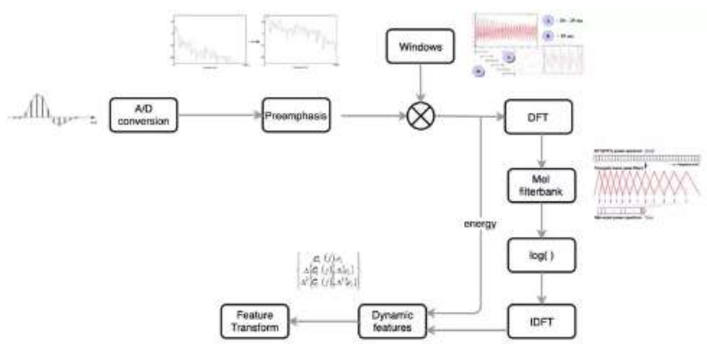
Thành phần chính

Mã nguồn Backend (Python):
- Flask, Flask-SocketIO, pydub, speech_recognition
- Xử lý chuyển đổi âm thanh, nhận dạng văn bản và xác định người nói

Mã nguồn Frontend (HTML, JavaScript):
- Giao diện hiển thị trạng thái cuộc họp, transcript, người nói
- Tích hợp Web Speech API và SocketIO để nhận dữ liệu từ server.

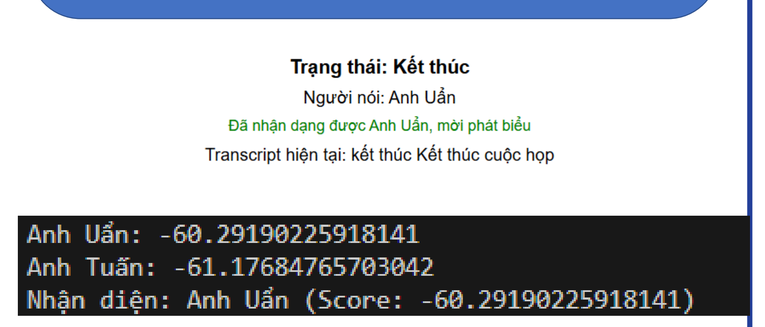


Mô hình nhận diện người nói Speaker Identification:
- Sử dụng mô hình học sâu được huấn luyện trước (pre-trained model) để trích xuất đặc trưng âm thanh (embedding) và so sánh với cơ sở dữ liệu người nói đã đăng ký.



Kết quả

Thực nghiệm:
- Hệ thống đã được kiểm tra trong các phiên họp trực tuyến, cho phép nhận diện và phân biệt người nói một cách hiệu quả
- Transcript tự động được lưu lại, hỗ trợ quá trình ghi chép và lưu trữ thông tin



Ứng dụng:
- Hỗ trợ tổ chức cuộc họp, phân tích nội dung trao đổi
- Phục vụ cho các ứng dụng giám sát và hỗ trợ điều khiển thiết bị qua giọng nói

Kết luận và các công trình trong tương lai

Kết Luận:
- Hệ thống chứng tỏ khả năng nhận diện giọng nói và xác định người nói theo thời gian thực, hỗ trợ việc lưu trữ transcript hiệu quả và mở rộng cho nhiều ứng dụng thực tế.

Hướng phát triển:
- Nâng cao độ chính xác nhận diện qua mô hình deep learning.Hỗ trợ đa ngôn ngữ và tích hợp thêm các chức năng phân tích nâng cao.Phát triển giao diện quản lý dữ liệu và báo cáo trực quan hơn.