NHẬN DIỆN GIỌNG NÓI

Xử Lý Tiếng Nói - INT3411_20

Thành Viên:
Đoàn Đình Dũng - 17021218
Hoàng Ngọc Dũng - 17021220

I. Giới Thiệu

Để đáp ứng cho các yêu cầu của bài tập cuối kỳ, chúng em đã lựa chọn chủ đề nhận dạng âm thanh và xây dựng lên một web app sử dụng Flask Python và đã tích hợp, áp dụng các chức năng xử lý âm thanh.

Các tính năng của web app này là sự kết hợp giữa nhận dạng giọng nói và tương tác máy. Người dùng có thể đưa file âm thanh dạng file .wav hoặc trực tiếp ghi âm lại giọng nói của mình và nhận lại kết quả hiển thị ở dạng văn bản. Ngoài ra việc nhận dạng giọng nói trực tiếp của người dùng cũng đã được chúng em mở rộng bằng việc kết hợp với tương tác máy. Người dùng khi sử dụng web app sẽ có thể dùng giọng nói của mình để mở ra các tab web mới như Google, Youtube hay Facebook. Hiện tại tất cả các chức năng đều thích hợp để sử dụng với ngôn ngữ là tiếng Việt.

II. Quá Trình Xây Dựng

1. Các Công Nghệ Sử Dụng

- **Python:**

Là một ngôn ngữ lập trình bậc cao với các ưu điểm cốt lõi như mạnh dễ đọc và áp dụng. Cấu trúc hoạt động của Python đề cao sự đơn giản do đó khiến người dùng có thể dễ dàng làm quen và sử dụng một cách hiệu quả. Đây cũng là một trong những ngôn ngữ lập trình phù hợp nhất cho các ứng dụng, công việc có liên quan đến học máy. Toàn bộ dự

án của bài tập này được xây dựng sử dụng Python cũng như các tính năng, package mở rộng của nó.

- **Thư viện Speech Recognition:**

Thư viện Speech Recognition của Python sẽ đảm nhiệm việc xử lý các dữ liệu phục vụ cho việc nhận dạng giọng nói với sự hỗ trợ cho một số công cụ và API. Thư viện này có thể hoạt động trong cả hai môi trường trực tuyến và ngoại tuyến. Chúng em quyết định sử dụng thư viện này vì Speech Recognition tương thích với nhiều loại API và có hỗ trợ cho ngôn ngữ tiếng Việt.

- **Flask:**

Flask là một Framework xây dựng giao diện web được viết bằng ngôn ngữ Python. Do không yêu cầu các công cụ hoặc thư viện cụ thể nên Flask được xếp vào loại micro-framework với các ưu điểm như nhẹ, độc lập, ít sử dụng các thư viện khác bên ngoài cũng như giúp người dùng có thể dễ dàng phát hiện, xử lý các lỗ hổng bảo mật.

Trong bài tập cuối kỳ này, bọn em đã sử dụng Flask để tạo nên giao diện cho web app cùng với các nút chức năng ví dụ như tải file lên web, bắt đầu ghi âm để nhận dạng giọng nói.

2. Bộ Dữ Liệu

Web app sử dụng bộ dữ liệu đã được tích hợp sẵn trong thư viện Speech Recognition của Python. Đây là bộ dữ liệu bao gồm các âm, từ, câu được tổng hợp và huấn luyện từ nhiều ngôn ngữ khác nhau. Điều này cho phép việc nhận diện âm thanh một cách tổng quan và tương đối chính xác. Do đó mà việc nhận dạng âm thanh của web app cũng hoạt động tốt với ngôn ngữ tiếng Việt.

3. Xây Dựng Các Chức Năng

Web app gồm hai chức năng chính:

- Đưa file ghi âm lên để xuất ra văn bản:

```python
if request.method == "POST":
        print("FORM DATA RECEIVED")

        if "file" not in request.files:
            return redirect(request.url)

        file = request.files["file"]
        if file.filename == "":
            return redirect(request.url)
        print('\n\nfile upoaded:', file.filename)

        if file:
            recognizer = sr.Recognizer()
            audioFile = sr.AudioFile(file)
            with audioFile as source:
                data = recognizer.record(source)
            transcript = recognizer.recognize_google(data, language="vi")
```

Đoạn code trên sẽ kiểm tra phương thức truyền dữ liệu, sau khi kiểm tra xong sẽ kiểm tra sự tồn tại của file đã được tải lên chưa và sau đó sẽ sử dụng thư viện Speech Recognition của Python để chuyển dữ liệu từ âm thanh sang văn bản và đưa lên trang web.

- Ghi âm trực tiếp trên web, ngoài ra khi xuất hiện một từ khóa có thể hiện ra phần tìm kiếm trên web của Youtube, Google hay Facebook.

```python
if request.method == "GET":
        r1 = sr.Recognizer()
        r2 = sr.Recognizer()
        r3 = sr.Recognizer()
        r4 = sr.Recognizer()

        with sr.Microphone() as source: #sử dụng đầu vào từ micro làm nguồn âm thanh
            print('[search video: search youtube]')
            print('speak now')
            audio = r3.listen(source) #nghe cụm từ đầu tiên rồi trích nó làm dữ liệu âm thanh

        if 'video' in r2.recognize_google(audio):
            r2 = sr.Recognizer()
            url = 'https://www.youtube.com/results?search_query='
            with sr.Microphone() as source:
                print('search your query')
                audio = r2.listen(source)

                try:
                    text = r2.recognize_google(audio, language='vi')
                    print(text)
                    wb.get().open_new(url + text)
                    print("You said : {}".format(text))
                except sr.UnknownValueError:
                    print('Error')
                except sr.RequestError as e:
                    print('failed'.format(e))
```

Đoạn code trên sẽ sử dụng để truyền dữ liệu âm thanh trực tiếp từ microphone vào và xử lý. Khi xử lý nếu dữ liệu được truyền vào có từ khóa như video, google, facebook thì sau đó sẽ tiếp nhận thêm dữ liệu âm thanh mới từ microphone và trực tiếp tìm kiếm, hiển thị dữ liệu trên web yêu cầu và song song đó nó cũng in ra đoạn văn bản mà mình tìm kiếm. Nếu không có từ khóa thì nó sẽ tự động tin ra văn bản.

4. Một Số Khó Khăn

Trong quá trình xây dựng web app này, bọn em nhận thấy việc ứng dụng sẽ gặp phải một số khó khăn thường thấy trong việc nhận dạng âm thanh và xử lý ngôn ngữ tự nhiên như:

- Mỗi người sẽ có một giọng điệu và các cách phát âm khác nhau điều này khiến việc nhận dạng các từ có thể gặp khó khăn.
- Tùy vào địa điểm thu âm mà có thể lẫn các tạp âm hoặc tiếng vang.
- Các ngôn ngữ khác nhau cũng sẽ ảnh hưởng đến âm điệu khi nói.

III. Kết Luận Và Mở Rộng

Hiện tại chúng em đã xây dựng được một web app nhận dạng âm thanh với những chức năng thiết yếu nhất. Web app này có thể sử dụng và những mục đích như trợ giúp con người trong tương tác máy bằng giọng nói hoặc giúp những người khiếm thính hiểu được nội dung của một đoạn âm thanh.

Hướng mở rộng của web app này có thể là tích hợp các chức năng khác như trợ lý ảo hoặc các chức năng sinh âm thanh. Ngoài ra việc sử dụng Speech Recognition cũng là tiền đề để xây dựng nên những hệ thống xác thực bằng giọng nói.