**Báo Cáo**

**1. Nhận dạng thực thể được đặt tên**

**Cách 1: Sử dụng spacy kết hợp rule-based**

A diagram of a computer

Description automatically generated

*Hình 1.1 biểu diễn thuật toán*

**Có 2 bước chính trong thuật toán như sau:**

*Bước 1: Trích xuất dữ liệu từ PDF*

Dữ liệu đầu vào là các tệp PDF, từ đó tôi sử dụng thư viện pdfplumber để trích xuất toàn bộ thông tin. Tuy nhiên, để tối ưu hóa tốc độ của thuật toán, tôi nhận thấy rằng tên và email thường xuất hiện trong khoảng 10 dòng đầu tiên và thường nằm trong 2 trang đầu tiên nếu CV dài hơn 2 trang. Ngược lại, nếu CV ngắn hơn 2 trang, ta sẽ lấy toàn bộ các trang của CV. Điều này giúp giảm thiểu dữ liệu cần xử lý, từ đó tăng tốc độ thuật toán.

*Bước 2: Trích xuất tên và email*

*Bước 2.1: Sử dụng Spacy*

Để trích xuất tên, tôi sử dụng mô hình nhận dạng thực thể có tên (NER) của Spacy với khả năng nhận diện các thực thể thuộc loại "PERSON". Tuy nhiên, kết quả nhận diện này chưa thật sự chính xác, vì Spacy có thể xác định các tên đường phố, địa điểm như là thực thể "PERSON".

*Bước 2.2: Sử dụng phương pháp dựa trên quy tắc*

Để cải thiện độ chính xác của việc nhận diện tên, tôi kết hợp thêm các quy tắc lọc. Cụ thể, tên phải có ít nhất hai từ và họ phải nằm trong danh sách các họ phổ biến của người Việt Nam. Khi Spacy tìm thấy các thực thể "PERSON" và các thực thể này thỏa mãn cả hai điều kiện trên, quá trình nhận diện tên sẽ dừng lại để tránh trùng lặp.

Đối với việc trích xuất email, tôi sử dụng biểu thức chính quy (regular expressions) để dễ dàng xác định và trích lọc các địa chỉ email từ văn bản.

Bằng cách kết hợp cả hai phương pháp này, tôi có thể trích xuất thông tin tên và email từ các tệp PDF một cách hiệu quả hơn.

**2. Text to speech**

**2.1. Giới thiệu**

Công nghệ chuyển đổi văn bản thành giọng nói (Text-to-Speech - TTS) là một lĩnh vực nghiên cứu quan trọng trong xử lý ngôn ngữ tự nhiên (NLP) và trí tuệ nhân tạo (AI). TTS có thể chuyển đổi văn bản thành âm thanh, cho phép máy tính đọc to tài liệu, website, và sách bằng các giọng nói tự nhiên. TTS mang lại lợi ích lớn cho những người cần truy cập nội dung văn bản dưới dạng âm thanh, và cung cấp cách giao tiếp bao hàm và dễ tiếp cận hơn cho nhiều người.

**2.2. Tổng quan về Pipeline Xây dựng Mô hình TTS cho Tiếng Việt**

Việc xây dựng mô hình TTS cho tiếng Việt đòi hỏi nhiều bước phức tạp để đảm bảo chất lượng giọng nói đầu ra là tự nhiên, dễ hiểu và chính xác về mặt ngữ điệu. Dưới đây là pipeline đề xuất:

*Bước 1: Thu thập và xử lý dữ liệu*

- Thu thập dữ liệu: Thu thập một lượng lớn dữ liệu giọng nói tự nhiên từ các diễn giả tiếng Việt. Các dữ liệu này bao gồm các đoạn hội thoại, đọc báo, sách, và các nguồn tài liệu khác với đa dạng ngữ điệu và tốc độ nói.

- Chuyển đổi văn bản: Các văn bản tương ứng với dữ liệu giọng nói sẽ được thu thập và chuẩn hóa. Đảm bảo văn bản phản ánh đúng ngữ cảnh của giọng nói và bao gồm các yếu tố như dấu câu, từ viết tắt, số, và ký hiệu.

- Gắn nhãn dữ liệu: Dữ liệu giọng nói cần được gắn nhãn để tương ứng với văn bản. Điều này bao gồm việc chia nhỏ đoạn âm thanh thành các câu hoặc cụm từ và gắn nhãn chúng với văn bản tương ứng.

*Bước 2: Xử lý văn bản (Text Processing)*

- Chuẩn hóa văn bản: Bao gồm xử lý ký tự đặc biệt, số, viết tắt và chuyển đổi chúng thành dạng mà mô hình có thể xử lý dễ dàng. Đối với tiếng Việt, cần đặc biệt chú ý đến dấu câu và các ký tự đặc biệt như dấu mũ và dấu nặng.

- Phân tích cú pháp:Phân tích cú pháp giúp xác định cấu trúc ngữ pháp của câu và hỗ trợ việc chuyển đổi từ văn bản sang giọng nói. Đây là bước quan trọng để đảm bảo giọng đọc có nhịp điệu và ngữ điệu tự nhiên.

*Bước 3: Xử lý âm thanh (Speech Processing)*

Mô hình âm vị học (Phonetic Modeling): Xây dựng mô hình âm vị học cho tiếng Việt để xác định cách phát âm các âm vị dựa trên văn bản đầu vào. Tiếng Việt có đặc thù về ngữ điệu và âm điệu, vì vậy mô hình cần được huấn luyện để nhận diện và xử lý tốt các yếu tố này.

Chuyển đổi văn bản thành âm vị (Text-to-Phoneme Conversion): Chuyển đổi văn bản sang chuỗi âm vị dựa trên mô hình âm vị học. Đây là bước quan trọng để mô hình hiểu rõ cách phát âm từng từ hoặc cụm từ trong ngữ cảnh.

*Bước 4: Tổng hợp giọng nói (Speech Synthesis)*

Mô hình tổng hợp giọng nói (Speech Synthesis Model):Mô hình tổng hợp giọng nói sẽ chuyển đổi chuỗi âm vị thành sóng âm. Có nhiều cách tiếp cận cho bước này, nhưng một trong những phương pháp hiện đại nhất là sử dụng mạng nơ-ron sâu, chẳng hạn như WaveNet, Tacotron 2, hoặc FastSpeech.

Điều chỉnh ngữ điệu và tốc độ (Prosody and Speed Adjustment): Ngữ điệu, nhịp điệu, và tốc độ của giọng nói cần được điều chỉnh để đảm bảo giọng đọc tự nhiên và dễ nghe. Điều này đặc biệt quan trọng với tiếng Việt, do ngữ điệu ảnh hưởng trực tiếp đến nghĩa của từ.

*Bước 5: Hậu xử lý (Post-Processing)*

Lọc và tinh chỉnh âm thanh: Sau khi tổng hợp giọng nói, bước cuối cùng là áp dụng các kỹ thuật lọc để loại bỏ tiếng ồn, điều chỉnh âm lượng, và tinh chỉnh chất lượng âm thanh.

Đánh giá chất lượng: Sử dụng các phương pháp đánh giá chủ quan và khách quan để đảm bảo giọng nói đầu ra đáp ứng được yêu cầu về chất lượng. Có thể sử dụng các thước đo như Mean Opinion Score (MOS) để đánh giá chất lượng tổng thể của giọng nói.3. Phân tích các thách thức và giải pháp

Thách thức 1: Thu thập và xử lý dữ liệu giọng nói

Vấn đề: Việc thu thập dữ liệu giọng nói đa dạng và đủ lớn để huấn luyện mô hình là thách thức lớn. Đặc biệt, cần đảm bảo rằng dữ liệu giọng nói phản ánh đầy đủ các vùng miền và ngữ điệu khác nhau trong tiếng Việt.

Giải pháp: Tìm kiếm và hợp tác với các tổ chức, đơn vị có sẵn dữ liệu giọng nói hoặc tiến hành các cuộc khảo sát, thu thập từ cộng đồng. Đảm bảo đa dạng về vùng miền và ngữ cảnh để dữ liệu đủ phong phú.

Thách thức 2: Xử lý đặc trưng ngữ âm tiếng Việt

Vấn đề: Tiếng Việt có nhiều đặc trưng ngữ âm độc đáo, đặc biệt là âm điệu (tonal language). Điều này làm cho việc chuyển đổi âm vị và tổng hợp giọng nói trở nên phức tạp.

Giải pháp: Xây dựng một mô hình âm vị học chuyên biệt cho tiếng Việt và sử dụng các kỹ thuật học sâu để học các mẫu ngữ âm từ dữ liệu giọng nói.

Thách thức 3: Đảm bảo ngữ điệu và tốc độ tự nhiên

Vấn đề: Ngữ điệu và tốc độ của giọng nói đầu ra phải tự nhiên để người nghe có thể dễ dàng hiểu và cảm nhận được nội dung.

Giải pháp: Sử dụng mô hình học sâu kết hợp với các kỹ thuật điều chỉnh ngữ điệu và tốc độ, như điều chỉnh nhịp điệu và độ nhấn trong giọng nói, để mô phỏng giọng đọc tự nhiên.