1. Giới thiệu
   1. Giới thiệu đề tài
   2. Mục tiêu và phạm vi đề bài

Tầm vực của đề tài là nghiên cứu, xây dựng và phát triển một ứng dụng di động cho phép nhân viên tạo báo cáo nhanh thông qua giọng nói. Mục tiêu của đề tài là chúng tôi sẽ hoàn thành xây dựng ứng dụng hỗ trợ và đảm bảo những chức năng cơ bản của ứng dụng tạo báo cáo nhanh bằng giọng nói. Ngoài ra, nhằm đề cao sự tương tác giữa nhân viên và người quản lý, ứng dụng sẽ được tích hợp vào một hệ thống cho phép nhân viên có thể giao nộp báo cáo trực tuyến cho người quản lý.

1. Phân tích vấn đề và hướng giải quyết
   1. Nhận diện giọng nói
      1. Vấn đề

Nhận dạng giọng nói là một quá trình nhận dạng mẫu, với mục đích là phân lớp (classify) thông tin đầu vào là tín hiệu tiếng nói thành một dãy tuần tự các mẫu đã được học trước đó và lưu trữ trong bộ nhớ. Các mẫu là các đơn vị nhận dạng, chúng có thể là các từ, hoặc các âm vị. Nếu các mẫu này là bất biến và không thay đổi thì công việc nhận dạng giọng nói trở nên đơn giản bằng cách so sánh dữ liệu giọng nói cần nhận dạng với các mẫu đã được học và lưu trữ trong bộ nhớ. Khó khăn cơ bản của nhận dạng giọng nói đó là giọng nói luôn biến thiên theo thời gian và có sự khác biệt lớn giữa giọng nói của những người nói khác nhau, tốc độ nói, ngữ cảnh và môi trường âm học khác nhau. Xác định những thông tin biến thiên nào của giọng nói là có ích và những thông tin nào là không có ích đối với nhận dạng giọng nói là rất quan trọng.

Thoạt nhìn thì việc triển khai công nghệ nhận dạng giọng nói khá đơn giản, nhưng thực chất thì không phải như thế. Thứ nhất, chúng ta cần phải xây dựng nên một công nghệ có thể lắng nghe, phân tích và phiên dịch một cách chính xác giọng nói của người dùng. Công nghệ này yêu cầu độ chính xác cao, để thiết bị có thể hiểu và phiên dịch chính xác những gì người dung nói, tránh gây ức chế cho người sử dụng. Có một kĩ thuật được nhắc đến nhiều trong thời gian gần đây, đó là Xử lý ngôn ngữ tự nhiên (Natural Language Processing - NLP). Nó là tập hợp của nhiều thuận toán phức tạp nhằm phân tích mệnh lệnh của người dùng nhưng không bắt buộc họ phải nói theo một cấu trúc câu định sẵn. Tuy nhiên NLP cũng không phải là đơn giản để phát triển. Cả Apple, Google và Microsoft đều phải đầu tư rất nhiều tiền bạc và nguồn lực để có thể đưa NLP lên đến mức tiến bộ như hiện nay. Mặc dù vậy, các công ty vẫn phải tiếp tục nghiên cứu để cải thiện độ chính xác.

Bên cạnh đó, vấn đề bản địa hóa (localization) cũng là một chuyện làm đau đầu các lập trình viên. Mỗi quốc gia sẽ có ngôn ngữ của riêng mình, vấn đề đó là làm thế nào để có thể hỗ trợ càng nhiều ngôn ngữ càng tốt.

* + 1. Giải pháp

Bởi vì việc nghiên cứu về nhận dạng giọng nói cũng như công nghệ xử lý ngôn ngữ tự nhiên đòi hỏi việc đầu tư nhiều về thời gian và nguồn lực nên nhóm quyết định sử dụng các thư viện và các api có sẵn của Google, Apple, Microsoft để phục vụ cho việc nhận dạng giọng nói trong ứng dụng.

* 1. Độ chính xác của các api
     1. Vấn đề

Công nghệ nhận diện giọng nói hiện nay của Google, Apple, Microsoft rất phát triển. Tuy nhiên, hiện nay công nghệ giọng nói vẫn chưa thể thay thế hoàn toàn bàn phím ảo/vật lý hoặc các nút trên màn hình. Các công ty vẫn phải tiếp tục nghiên cứu để cải thiện độ chính xác. Vì thế, vấn đề về độ chính xác của các api mà nhóm sử dụng cũng được đặt ra : Cần xử lý các đoạn văn bản được trả về từ các api.

* + 1. Giải pháp

Với vấn đề trên, nhóm quyết định xử dụng thêm các thư viện xử lý ngôn ngữ tự nhiên hiện có để xử lý các đoạn văn bản trả về từ api cho phù hợp với cấu trúc của một bài báo cáo. Nghĩa là sử dụng thư viện để xử lý đoạn văn bản bao gồm tiêu đề của từng trường và dữ liệu trả về tương ứng.

Bên cạnh đó, ứng dụng cũng cần cho phép người sử dụng chỉnh sửa những lỗi sai bằng bàn phím ảo của các thiết bị di động.

* 1. Âm điệu, âm vực khác nhau.
     1. Vấn đề

Giọng nói của mỗi người có âm điệu, âm vực khác nhau và biến thiên theo thời gian, không gian nên việc nhận dạng giọng nói không đơn thuần là việc so trùng các mẫu lưu trữ với các mẫu đầu vào.

* + 1. Giải pháp

Các api nhận diện giọng nói hiện nay có khả năng phân biệt âm điệu, âm vực khá ổn nên nhóm chọn dựa vào api để giải quyết vấn đề này.

* 1. Xử lí tiếng ồn của dữ liệu đầu vào
     1. Vấn đề

Trong hiện thực, người sử dụng dùng các thiết bị di động để báo cáo nhanh trong môi trường có nhiều tiếng ồn. Việc này sẽ gây ảnh hưởng tới việc nhập dữ liệu đầu vào bằng giọng nói nên âm thanh đầu vào cần được xử lý lọc tạp âm để tăng độ chính xác cho ứng dụng.

* + 1. Giải pháp

Các api nhận diện giọng nói hiện nay có khả năng xử lý tiếng ồn nên nhóm chọn dựa vào api để giải quyết vấn đề này.

* 1. Quản lý dữ liệu, người dùng
     1. Vấn đề

Ứng dụng cần là mối liên kết giữa nhân viên và cấp trên trong việc báo cáo kết quả của các cuộc họp. Vì thế, dữ liệu về các bản báo cáo cần được quản lý rõ ràng. Bên cạnh đó, các bản báo cáo cần được chuyển đến cấp trên ngay khi nó được tạo xong để được kiểm duyệt. Cùng với đó, cấp trên cũng cần biết được người tạo báo cáo, thời gian tạo và địa điểm để kiểm tra, quản lý nhân viên của mình.

* + 1. Giải pháp

Với vấn đề về việc quản lý, nhóm quyết định sử dụng server firebase để lưu trữ dữ liệu. Ngoài ra, bên cạnh ứng dụng di động dành cho nhân viên để tạo báo cáo, nhóm sẽ xây dựng thêm một trang web dành cho người quản lý để quản lý các bản báo cáo, nhân viên, thống kê các thông số. Trang web và ứng dụng sẽ sử dụng cùng một server để đảm bảo sự đồng bộ dữ liệu của các bản báo cáo. Ứng dụng sẽ yêu cầu thiết bị kết nối với mạng để gửi các bản báo cáo lên server. Cùng với đó, các bản báo cáo đều có yêu cầu về người tạo, thời gian và địa điểm tạo. Những thông số này sẽ được tạo cách tự động khi báo cáo được tạo.

* 1. Ngôn ngữ hỗ trợ
  2. Lưu trữ báo cáo
  3. Nhận diện dấu chấm, dấu phẩy
  4. Nhận diện chữ số
  5. Các thông số quan tâm
  6. Bảo mật thông tin
  7. Lưu trữ dữ liệu
  8. Kết nối mạng

1. Các ứng dụng, công nghệ hiện tại
   1. Google Speech To Text API
   2. IBM Watson Speech To Text API
   3. Microsoft Speech API
   4. Một số ứng dụng sử dụng công nghệ Speech To Text
2. Phân tích hệ thống
   1. Chức năng hệ thống
      1. Người dùng
      2. Người quản trị
   2. Use Case Diagram
      1. Ứng dụng di động tạo báo cáo nhanh bằng giọng nói
         1. Overview
         2. Usecase diagram
         3. Đặc tả usecase
      2. Web quản lý báo cáo
         1. Overview
         2. Usecase diagram
         3. Đặc tả usecase
   3. Class Diagram
3. Tổng kết
   1. Các kết quả đã đạt được
   2. Hướng phát triển

Tài liệu tham khảo