

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**  
**TRƯỜNG ĐẠI HỌC CẦN THƠ**  
**KHOA CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

---



**HUỲNH THANH TOÀN**

**CHATBOT TÌM KIẾM PHÁP LUẬT**

**NIÊN LUẬN NGÀNH KHOA HỌC MÁY TÍNH**

**Cần Thơ – 2020**

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**  
**TRƯỜNG ĐẠI HỌC CẦN THƠ**  
**KHOA CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

---



**CHATBOT TÌM KIẾM PHÁP LUẬT**

**NIÊN LUẬN NGÀNH KHOA HỌC MÁY TÍNH**

**CÁN BỘ HƯỚNG DẪN**

TS. Lưu Tiến Đạo

**SINH VIÊN THỰC HIỆN**

Huỳnh Thanh Toàn

MSSV: B1404453

**Cần Thơ – 2020**

[illegible]

## **LỜI CẢM ƠN**

Tôi xin bày tỏ lòng biết ơn chân thành đến những người sau đây, những người đã đã góp sức để luận văn của tôi được hoàn thành.

Đầu tiên tôi xin gửi lời cảm ơn đến gia đình tôi vì sự quan tâm và sự hỗ trợ của họ trong suốt những năm qua.

Xin cảm ơn thầy Lưu Tiến Đạo người đã tạo điều kiện cho tôi có thể thực hiện niên luận, cảm ơn thầy đã tận tình hướng dẫn, truyền đạt những kiến thức cho tôi.

Xin chân thành cảm ơn các Thầy Cô Giảng viên Đại học Cần Thơ, các thầy cô Khoa CNTT & TT nói chung và các thầy cô bộ môn Khoa Học Máy Tính nói riêng, những người đã truyền đạt những kiến thức quý báu trong thời gian qua.

Tôi thành thật biết ơn tất cả mọi người.

## DANH SÁCH HÌNH

| Hình | Tựa hình   | Trang |
|------|--|-------|
| 2.1  | Các bước hoạt động của hệ thống                      | 5     |
| 2.2  | Các bước xây dựng dữ liệu huấn luyện                 | 6     |
| 2.3  | Xây dựng ý định người dùng                           | 6     |
| 2.4  | Các câu trả lời của chatbot                          | 7     |
| 2.5  | Khung kịch bản của chatbot                           | 7     |
| 2.6  | Chuyển câu hỏi dạng âm thanh thành văn bản           | 8     |
| 2.7  | Chuyển câu trả lời văn bản thành giọng nói           | 8     |
| 2.8  | Phát âm thanh câu trả lời                            | 8     |
| 3.1  | Xếp hạng cho các ý định                              | 9     |
| 3.2  | Biểu đồ ma trận hỗn loạn dữ liệu ngôn ngữ tự nhiên   | 13    |
| 3.3  | Ma trận hỗn loạn dữ liệu các câu trả lời của chatbot | 14    |

## DANH SÁCH BẢNG

| <b>Bảng</b> | <b>Tên bảng</b>   | <b>Trang</b> |
|-------------|---|--------------|
| 3.1         | Số lần thử nghiệm và tỷ lệ trả lời đúng câu hỏi của chatbot | 9            |
| 3.2         | Các câu hỏi và câu trả lời của botchat                      | 10           |

## TỪ VIẾT TẮT

| Từ viết tắt | Nghĩa tường minh                  |
|-------------|-----------------------------------|
| API         | Application Programming Interface |
| BGDĐT       | Bộ giáo dục và đào tạo            |
| CNTT        | Công nghệ thông tin               |
| DM          | Dialogue Management               |
| NLU         | Natural Language Understanding    |
| NLG         | Natural Language Generation       |

## TÓM TẮT

Đề tài được thực hiện với mục đích xây dựng trợ lý ảo hỗ trợ tìm kiếm pháp luật hỗ trợ người sử dụng hỏi đáp và tư vấn về các vấn đề liên quan đến pháp luật thuộc phạm vi thông tư số 09/2017/TT-BGDĐT về việc quy định điều kiện, trình tự, thủ tục mở ngành hoặc chuyên ngành đào tạo và đình chỉ tuyển sinh, thủ tục hỏi quyết định mở ngành hoặc chuyển ngành đào tạo trình độ thạc sĩ, trình độ tiến sĩ. Đề tài sử dụng RASA framework để xây dựng mô hình chatbot và sử dụng các thư viện Pyaudio, Speech Recognition của python để thu âm qua microphone và chuyển giọng nói thành văn bản qua các API của Google và FPT tạo tính năng tương tác với người sử dụng bằng giọng nói.

Kết quả đề tài đã xây dựng được bộ dữ liệu bao gồm câu hỏi ngôn ngữ tự nhiên phục vụ quá trình huấn luyện chatbot với 163 câu hỏi mẫu và 25 câu trả lời tương ứng trong phạm vi “Thông tư số 09/2017/TT-BGDĐT”. Qua 3 lần kiểm thử với 12 câu hỏi ngẫu nhiên nằm trong kịch bản thu được kết quả 11/10, độ chính xác của bot hiểu được câu hỏi là 91,6% và độ chính xác khi bot phát sinh câu trả lời là 97,3% và có thể tương tác với người sử dụng bằng giọng nói.



## MỤC LỤC

|   |     |
|---|-----|
| LỜI CẢM ƠN.....                             | i   |
| DANH SÁCH HÌNH.....                         | ii  |
| DANH SÁCH BẢNG.....                         | iii |
| TỪ VIẾT TẮT.....                            | iv  |
| TÓM TẮT .....                               | v   |
| MỤC LỤC.....                                | vi  |
| PHẦN GIỚI THIỆU.....                        | 1   |
| 1. Đặt vấn đề .....                         | 1   |
| 2. Lịch sử giải quyết vấn đề.....           | 1   |
| 3. Mục tiêu .....                           | 2   |
| 3.1 Mục tiêu chung.....                     | 2   |
| 3.2 Mục tiêu cụ thể.....                    | 2   |
| 4. Đối tượng nghiên cứu .....               | 2   |
| 5. Phương pháp nghiên cứu .....             | 2   |
| 6. Bố cục niên luận .....                   | 3   |
| PHẦN NỘI DUNG.....                          | 4   |
| CHƯƠNG 1: MÔ TẢ BÀI TOÁN.....               | 4   |
| 1.1 Mô tả chi tiết bài toán .....           | 4   |
| 1.2 Các vấn đề và giải pháp liên quan ..... | 4   |
| 1.2.1 Vấn đề .....                          | 4   |
| 1.2.2 Giải pháp .....                       | 4   |
| CHƯƠNG 2: THIẾT KẾ VÀ CÀI ĐẶT.....          | 5   |
| 2.1 Thiết kế .....                          | 5   |
| 2.2 Xây dựng hệ thống.....                  | 5   |
| 2.2.1 Ứng dụng RASA xây dựng chatbot .....  | 5   |
| 2.2.2 Xử lý nhập xuất bằng giọng nói.....   | 8   |
| CHƯƠNG 3: THỬ NGHIỆM VÀ ĐÁNH GIÁ.....       | 10  |
| 3.1 Thử nghiệm.....                         | 10  |
| 3.2 Đánh giá.....                           | 13  |
| PHẦN KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....      | 15  |
| 1. Kết luận.....                            | 15  |
| 2. Hướng phát triển .....                   | 15  |
| TÀI LIỆU THAM KHẢO .....                    | 16  |

# PHẦN GIỚI THIỆU

## 1. Đặt vấn đề

Trước kia, con người tiếp cận thông tin hằng ngày thông qua sách báo, tạp chí hoặc các thư viện được đặt tại các trung tâm của các thành phố lớn, điều này là khá khó khăn đối với những người ở những vùng ngoại ô hay không thể mua sách, báo hằng ngày. Khi internet du nhập vào Việt Nam năm 1997 đánh dấu một sự chuyển biến mới cùng với sự phát triển không ngừng của công nghệ thông tin các loại sách báo và tạp chí cũng dần được số hóa để đưa thông tin, tin tức tiếp cận được với người dân dễ dàng hơn, bên cạnh đó là những mạng xã hội như Facebook, Twitter, Lotus, ... các thông tin như văn hóa, giải trí, giáo dục, thể thao, xã hội,... được cập nhật hàng ngày hàng giờ. Nhưng vẫn có những loại thông tin rất khó có thể tiếp cận với người dân điển hình đó là pháp luật, những nghị định, thông tư và vô số những điều, khoản mà đa số ít người đọc đến.

Chatbot hay còn gọi là trợ lý ảo là một lĩnh vực khá nổi trội trong những năm trở lại đây, với cách thức người sử dụng đặt câu hỏi hoặc yêu cầu và máy tính trả lời hoặc thực hiện. Là công cụ hỗ trợ đắc lực dành cho các công ty, tổ chức trong việc phát triển, duy trì và cải thiện mối quan hệ với khách hàng.

Từ những vấn đề trên có thể thấy đề tài “*chatbot tìm kiếm pháp luật*” là hết sức cần thiết, nhằm tạo ra một trợ lý ảo có thể giải đáp những thắc mắc về pháp luật, giúp cho người sử dụng có thể nắm bắt một cách chi tiết nhất về pháp luật.

## 2. Lịch sử giải quyết vấn đề

Đã có rất nhiều đề tài nghiên cứu trong lĩnh vực chatbot điển hình như:

“Chatbot cho sinh viên công nghệ thông tin”(Đỗ Thanh Nghị, 2019) đã xây dựng Chatbot trên máy tính Raspberry Pi có thể trả lời tự động cho sinh viên ngành Công nghệ thông tin (CNTT) các câu hỏi liên quan đến môi trường học tập và phương pháp học tập bậc đại học, kỹ năng nghề nghiệp, xu hướng công nghệ. Sinh viên thực hiện đặt câu hỏi cho Chatbot bằng cách nói trực tiếp bằng tiếng nói, Chatbot tiếp nhận câu hỏi từ microphone và chuyển đổi câu hỏi từ dạng tiếng nói sang dạng văn bản, nhận dạng câu hỏi văn bản bằng mô hình phân lớp văn bản đã được huấn luyện, Chatbot trả về câu trả lời dạng văn bản và phát ra câu trả lời bằng tiếng nói cho sinh viên nghe.

Hay như nghiên cứu “Một chatbot mới cho dịch vụ khách hàng trên phương tiện truyền thông” (Anbang Xu và ctv, 2017) đã tạo ra một hệ thống đàm thoại mới có thể tự động tạo phản hồi cho các yêu cầu của người dùng trên phương tiện truyền thông và mạng xã hội. Hệ thống được tích hợp các kỹ thuật deep learning và được đào tạo từ dữ liệu gần 1 triệu cuộc hội thoại trên Twitter giữa người dùng và các doanh nghiệp.

Đã có rất nhiều đề tài nghiên cứu về chatbot với nhiều lĩnh vực khác nhau như tư vấn tuyển sinh, ngân hàng, mạng xã hội,... nhưng về lĩnh vực pháp luật thì tương đối ít hoặc không có những nghiên cứu về chatbot để hỗ trợ, tư vấn cho người sử dụng về pháp luật. Do đó đề tài “*chatbot tìm kiếm pháp luật*” được thực hiện nhằm tạo ra một trợ lý ảo có thể tư vấn về mặt pháp luật và góp phần tư liệu cho những nghiên cứu chatbot sau này.

### **3. Mục tiêu**

#### *3.1 Mục tiêu chung*

- Xử lý câu hỏi của người dùng và câu trả lời của bot chat bằng giọng nói.
- Bot có thể hiểu được câu hỏi của người dùng và đưa ra câu trả lời chính xác nhất.

#### *3.2 Mục tiêu cụ thể*

- Nhận diện giọng nói của người dùng qua microphone và chuyển thành văn bản.
- Xây dựng tập tin nlu bao gồm các câu hỏi mà người dùng có thể hỏi.
- Xây dựng domain bao gồm các câu trả lời tương ứng với những câu hỏi của người dùng.
- Xây dựng cốt truyện cho bot chat có thể giao tiếp với người sử dụng.
- Chuyển đổi câu trả lời của bot từ văn bản thành giọng nói.

### **4. Đối tượng nghiên cứu**

Đối tượng nghiên cứu: người sử dụng có các câu hỏi liên quan đến pháp luật.

Phạm vi nghiên cứu: Thông tư số 09/2017/TT-BGDĐT về việc quy định điều kiện, trình tự, thủ tục mở ngành hoặc chuyên ngành đào tạo và đình chỉ tuyển sinh, thu hồi quyết định mở ngành hoặc chuyên ngành đào tạo trình độ thạc sĩ, trình độ tiến sĩ.

### **5. Các bước thực hiện**

B1: Xây dựng các dữ liệu cần thiết

nlu.md: là tập tin ngôn ngữ tự nhiên gồm các câu hỏi mà người sử dụng có thể hỏi, được định nghĩa trong các (intent – “ý đồ người dùng”).

domain.yml: định nghĩa các câu phản hồi của bot với các ý đồ người dùng gọi là các utter ứng với các intent của nlu.md.

stories.md: xây dựng các câu trò chuyện khi người dùng hỏi và bot trả lời.

B2: Train model

NLU: vector hóa ngôn ngữ, phân loại ý định người dùng.

DM: dựa vào trạng thái ngữ cảnh hội thoại để xác định ra các action xử lý cho câu đầu vào.

NLG: bot sinh câu trả lời dựa vào dữ liệu thành phần DM theo các mẫu câu template đã được xây dựng.

B3: Xử lý nhập xuất

Đầu vào: ghi nhận câu hỏi và chuyển đổi giọng nói thành văn bản để bot hiểu được câu hỏi và phát sinh câu trả lời.

Đầu ra: chuyển đổi câu trả lời mà bot phát sinh thành giọng nói.

## **6. Bố cục niên luận**

Phần giới thiệu

Giới thiệu tổng quát về đề tài.

Phần nội dung

Chương 1 : Mô tả bài toán

Chương 2 : Thiết kế, cài đặt giải thuật, biểu diễn cơ sở dữ liệu, trình bày các bước xây dựng hệ thống

Chương 3 : Kiểm thử hệ thống và đánh giá độ chính xác của hệ thống.

Phần kết luận

Trình bày kết quả đạt được và hướng phát triển hệ thống.

# PHẦN NỘI DUNG

## CHƯƠNG 1: MÔ TẢ BÀI TOÁN

### 1.1 Mô tả chi tiết bài toán

Phần quan trọng nhất của một chatbot đó là hiểu được ngôn ngữ tự nhiên, từ đó đưa ra câu trả lời cho người sử dụng. Với RASA framework câu hỏi dạng văn bản đầu vào được phân tách khoảng trắng thành các từ đơn nhận dạng từ ghép và từ đồng nghĩa. Sau đó tiến hành phân loại ý định dựa trên kỹ thuật starspace(Ledell Wu - 2017), xếp hạng cho các ý định và phát sinh câu trả lời với ý định tại hạng cao nhất. Câu trả lời là dạng văn bản được gửi đến FPT API tiến hành chuyển câu trả lời thành dạng âm thanh.

### 1.2 Các vấn đề và giải pháp liên quan

#### 1.2.1 Vấn đề

Các vấn đề chính để xây dựng chatbot bao gồm:

- Lưu câu hỏi của người dùng dưới dạng giọng nói.
- Chuyển câu hỏi thành văn bản để xử lý.
- Hiểu được câu hỏi của người dùng và đưa ra câu trả lời chính xác.

#### 1.2.2 Giải pháp

Để ghi âm câu hỏi của người dùng, đề tài sử dụng Pyaudio – thư viện Python cho phép kết nối đến microphone của máy tính và lưu âm thanh được nói vào microphone.

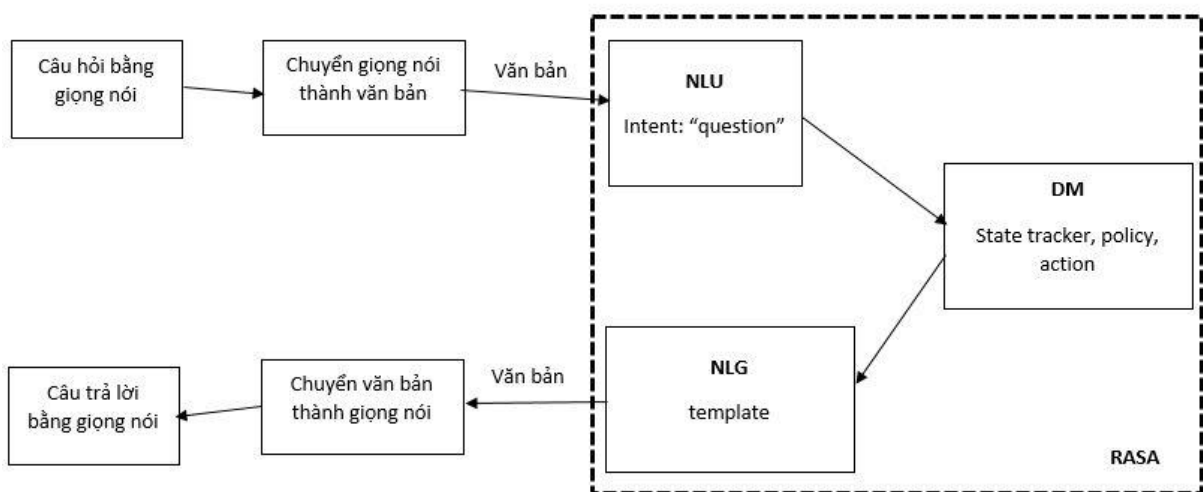
Để chuyển giọng nói thành văn bản, đề tài sử dụng Speech Recognition python library, thư viện này chứa các hàm cho phép kết nối đến Google voice API để xử lý âm thanh và chuyển âm thanh thành văn bản.

Sử dụng RASA framework để xử lý vấn đề ngôn ngữ tự nhiên và phát sinh câu trả lời.

## CHƯƠNG 2: THIẾT KẾ VÀ CÀI ĐẶT

### 2.1 Thiết kế

Người dùng đặt câu hỏi thông qua microphone trên máy tính, bot gửi dữ liệu câu hỏi dạng âm thanh đến Google API chuyển đổi giọng nói thành văn bản. Bot tiếp nhận câu hỏi dạng văn bản và bắt đầu phân loại ý định(intent) cho câu hỏi, xếp hạng cho các intent và chọn ra intent có hạng cao nhất. Khi đã hiểu được câu hỏi của người dùng bot phát sinh câu trả lời ứng với intent đó theo một kịch bản đã cho. Câu hỏi phát sinh ra ở dạng văn bản được gửi đến FPT API chuyển câu trả lời dạng âm thanh. Cuối cùng là tải là tải câu trả lời dạng âm thanh về máy và phát lên cho người dùng nghe.



Hình 2.1: Các bước hoạt động của hệ thống

### 2.2 Xây dựng hệ thống

#### 2.2.1 Ứng dụng RASA xây dựng chatbot

Rasa cung cấp cho ta 2 phương pháp chính xây dựng dữ liệu training cho bot:

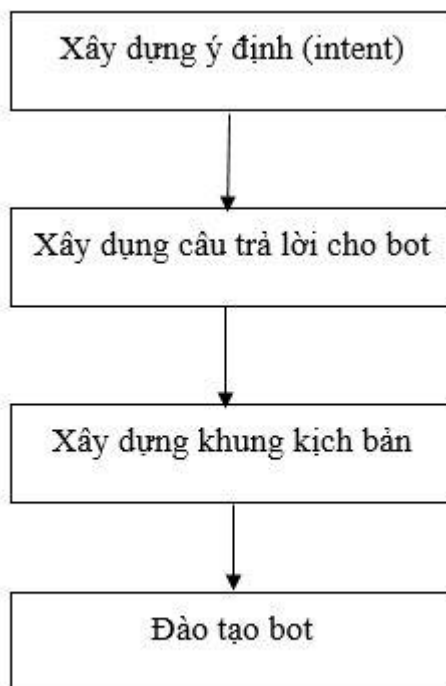
**Pretrained Embeddings (Intent\_classifier\_sklearn) :** Việc phân loại ý định người dùng sẽ dựa trên các tập dữ liệu được lọc trước, sau đó được sử dụng để thể hiện từng từ trong thông điệp người dùng dưới dạng từ nhúng (word embedding) hay biểu diễn ngôn ngữ dưới dạng vector(word2vec). Các tập dữ liệu này có thể được cung cấp từ Spacy hoặc MITIE ...

**Supervised Embeddings (Intent\_classifier\_tensorflow\_embedding):** Nhúng được giám sát. Với phương pháp này thì người dùng sẽ phải tự xây dựng dữ liệu từ đầu do ko có

dữ liệu đào tạo sẵn có. Nhưng với các bài toán trong một lĩnh vực nhỏ thì nó sẽ đảm bảo tính chính xác hơn nhiều và tránh dư thừa dữ liệu so với phương pháp ở trên.

Với bài toán tập trung vào miền đóng như trợ lý ảo hỗ trợ tìm kiếm pháp luật, được giới hạn trong phạm vi thì không cần sử dụng tập dữ liệu đào tạo từ trước (pretrained word embeddings) thay vào đó thì ta sẽ tự tạo tập dữ liệu training riêng của mình. Điều này cũng đảm bảo bot có thời gian training ngắn mà độ chính xác cao hơn.

Các bước chính để xây dựng chatbot (hình 2.2)



Hình 2.2: Các bước xây dựng dữ liệu huấn luyện

### *Xây dựng ý định*

Việc xây dựng ý định sẽ theo nguyên tắc là những mẫu câu hỏi của người dùng phổ biến nhất, thông dụng nhất có thể. Hiện tại bot đang hỗ trợ khoảng 24 intent (hình 2.3).

## intent:ask\_thacsi  
- điều kiện để mở ngành đào tạo thạc sĩ  
- làm sao để mở ngành thạc sĩ  
- em có biết làm sao để mở ngành thạc sĩ không  
- để mở ngành thạc sĩ thì cần gì  
- làm như thế nào để mở ngành thạc sĩ

### *Xây dựng câu trả lời cho bot*

Hình 2.3: Xây dựng ý định người dùng

Ứng với mỗi câu hỏi của người dùng thì ta cũng phải xây dựng các mẫu câu (template) trả lời của bot tương ứng (hình 2.4).

templates:

# Bot trả lời

utter\_thank:

- text: "không có gì, đó là trách nhiệm của tôi!"

utter\_unclear:

- text: "Câu hỏi của bạn có thể đã nằm ngoài nội dung. Hãy hỏi về những vấn đề mở ngành, quyết định mở ngành, đình chỉ hoặc thu hồi quyết định mở ngành, quy trình làm hồ sơ hay các vấn đề về thẩm định chất lượng"

utter\_thacsi:

- text: 'để mở ngành thạc sĩ cần đảm bảo các điều kiện: ngành đăng ký đào tạo, đội ngũ giảng viên, cơ sở vật chất và một số điều kiện khác.'

utter\_thacsi\_nganhdaotao:

- text: 'ngành đào tạo phải phù hợp với nhu cầu của người học và nhu cầu nguồn nhân lực cho phát triển kinh tế, xã hội của địa phương, vùng, cả nước. Tên ngành đào tạo phải có trong danh mục giáo dục. Ngành đào tạo trình độ đại học là ngành đúng hoặc gần đúng'

*Hình 2.4: Các câu trả lời của chatbot*

### *Xây dựng khung kịch bản*

## hỏi mở ngành thạc sĩ

\* ask\_thacsi

- utter\_thacsi

##hỏi ngành đào tạo

\* ask\_thacsi\_nganhdaotao

-

utter\_thacsi\_nganhdaotao

## hỏi đội ngũ giảng viên

\* ask\_thacsi\_doingugv

- utter\_thacsi\_doingugv

*Hình 2.5: Khung kịch bản của chatbot*



Dựa vào các tập ý định người dùng và các mẫu câu trả lời tương ứng của bot ta xây dựng khung kịch bản cho bot thông qua việc sắp xếp chung thành một đoạn hội thoại như (hình 2.5).

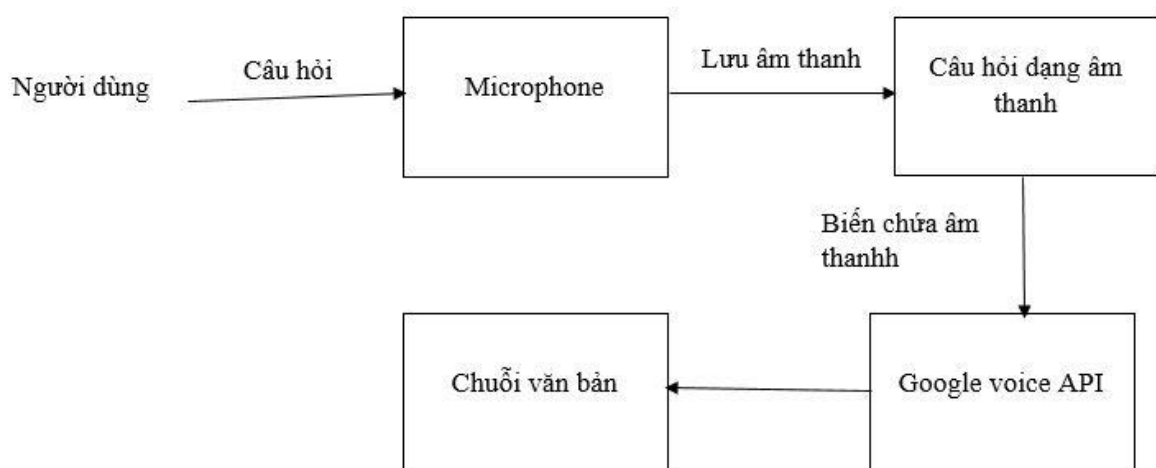
### 2.2.2 Xử lý nhập xuất bằng giọng nói

#### *Ghi âm chuyển đổi giọng nói thành văn bản*

Sau khi người dùng đọc câu hỏi thông qua microphone trên máy tính. Câu hỏi dạng âm thanh được lưu vào một biến tự định nghĩa nhờ sử dụng thư viện Pyaudio. Sau đó, kết nối đến Google voice API thông qua thư viện Speech Recognition để chuyển giọng nói thành văn bản. Do phải gửi yêu cầu chuyển âm thanh thành văn bản nên cần kết nối Internet để thực hiện

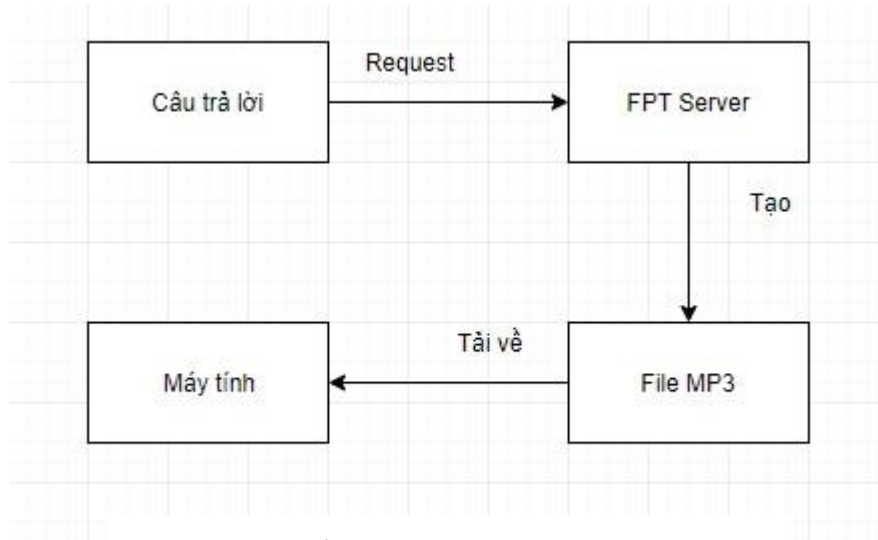
#### *Chuyển câu trả lời dạng văn bản thành giọng nói*

Đầu tiên thực hiện gửi gói tin request đến FPT server chứa các thông tin như giọng đọc, nội dung, tốc độ đọc. Sau khi chờ FPT server tạo file mp3 chứa nội dung vừa gửi trong gói tin đã được chuyển đổi thành giọng nói, thực hiện tải file này về máy tính bằng cách sử dụng thư viện Urllib.request của Python (hình 2.6).



Hình 2.6: Chuyển câu hỏi dạng âm thanh thành văn bản

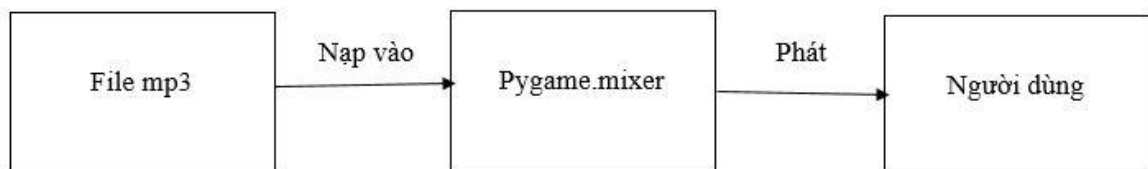
### *Phát file âm thanh chứa câu trả lời*



*Hình 2.7: Chuyển câu trả lời văn bản thành giọng nói*

Câu trả lời dạng văn bản sẽ được gửi đến máy chủ FPT xử lý và tạo thành tập tin âm thanh với phần mở rộng mp3, sau đó tiến hành tải tập tin vừa được tạo về máy tính (hình 2.7).

Sử dụng thư viện Pygame của python để thực hiện phát file âm thanh cho người sử dụng nghe dưới nền hệ thống (hình 2.8).



*Hình 2.8: phát âm thanh câu trả lời*

## CHƯƠNG 3: THỬ NGHIỆM VÀ ĐÁNH GIÁ

### 3.1 Thử nghiệm

Đề tài đã tiến hành thử nghiệm hỏi bot 10 câu hỏi ngẫu nhiên thu được kết quả ở bảng sau:

Bảng 3.1: Số lần thử nghiệm và tỷ lệ trả lời đúng câu hỏi của chatbot

| Lần thử nghiệm | Số câu trả lời đúng | Tỷ lệ trả lời đúng(%) |
|----------------|---------------------|-----------------------|
| 1              | 10/12               | 83%                   |
| 2              | 11/12               | 91,6%                 |
| 3              | 11/12               | 91,6%                 |

Thử nghiệm tương tác với bot 10 câu hỏi cho kết quả cả 3 lần đều là 100% và chỉ xoay quanh những câu hỏi trong kịch bản đã được đào tạo cho bot. Đối với những câu hỏi nằm ngoài phạm vi hoặc những câu hỏi không liên quan thì bot sẽ đưa ra câu trả lời gợi ý để điều hướng người dùng tới những câu hỏi mà bot có thể trả lời.

```
{'intent': {'name': 'ask_tiens', 'confidence': 0.4026506302386218}, 'entities': [], 'intent_ranking':  
[{'name': 'ask_tiens', 'confidence': 0.4026506302386218}, {'name': 'ask_thacsi', 'confidence':  
0.08760498525258921}, {'name': 'ask_dinhchi_xoa', 'confidence': 0.056765851148743905}, {'name':  
'ask_trinhthu', 'confidence': 0.04588399722955838}, {'name': 'ask_thutuc', 'confidence':  
0.045351148690881214}, {'name': 'ask_thuhoi', 'confidence': 0.037504648553012825}, {'name':  
'ask_quyetdinhmonganh', 'confidence': 0.0364055775279382}, {'name': 'ask_thamdinh',  
'confidence': 0.030984521500731967}, {'name': 'ask_thamquyen', 'confidence':  
0.022469443535085752}, {'name': 'ask_thacsi_cosovatchat', 'confidence': 0.022021735997442192}],  
'text': 'làm sao để mở ngành tiến sĩ'}
```

Hình 3.1: Xếp hạng cho các ý định

Khi nhận vào một câu hỏi bot sẽ tiến hành xếp hạng cho câu hỏi đó với các intent tương ứng và đưa ra câu trả lời ứng với intent có hạng cao nhất.

Bảng 3.2: Các câu hỏi và câu trả lời của botchat

| STT | Câu hỏi | Bot trả lời |
|-----|---------|-------------|
|-----|---------|-------------|

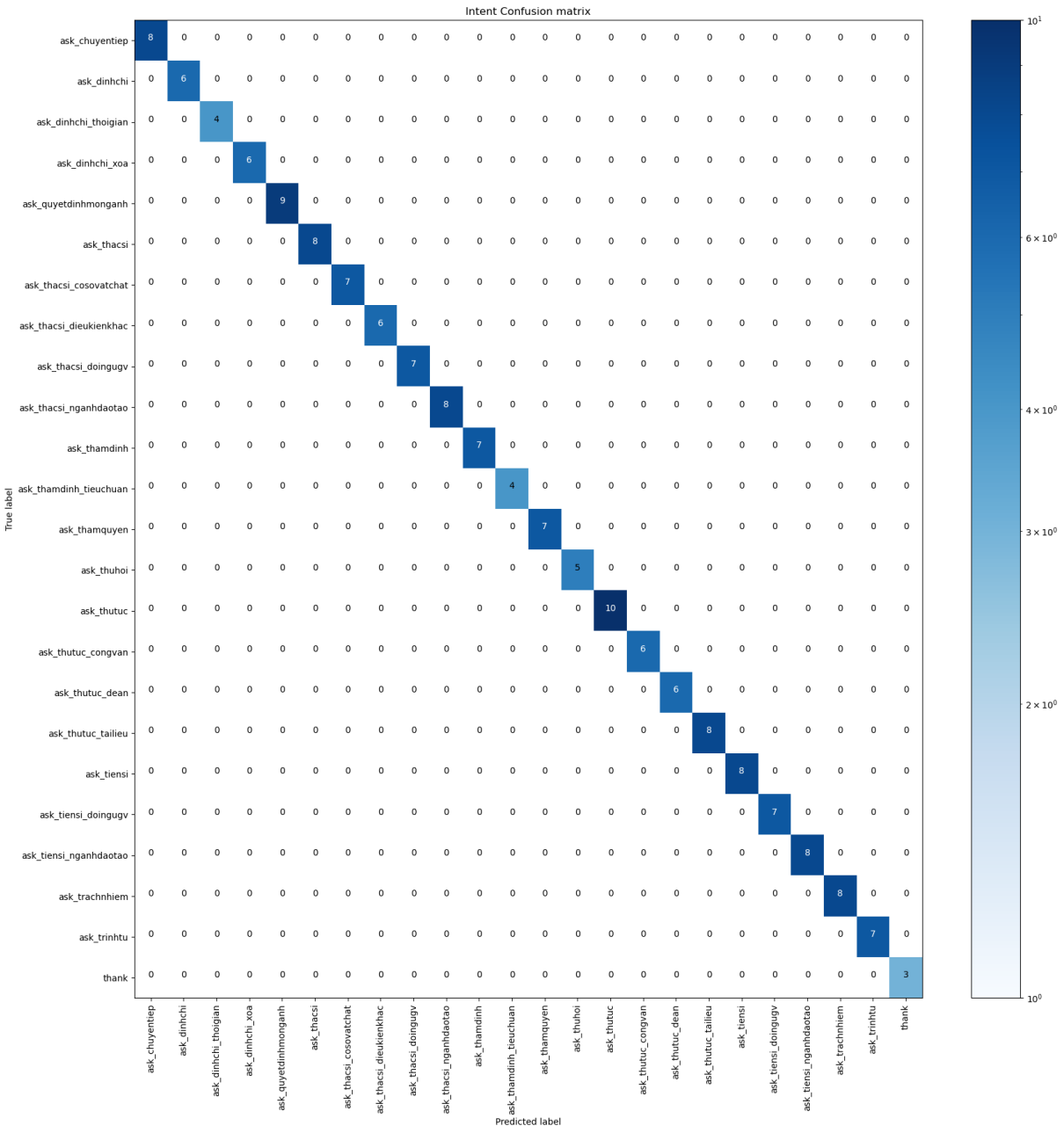
|   |   |   |
|---|---|---|
| 1 | làm cách nào để mở được ngành tiến sĩ                       | để mở ngành tiến sĩ cần đảm bảo các điều kiện: ngành đăng ký đào tạo, đội ngũ giảng viên, cơ sở vật chất và một số điều kiện khác   |
| 2 | đội ngũ giảng viên ngành tiến sĩ thì cần những điều kiện gì | có ít nhất 01 giáo sư và 03 tiến sĩ khoa học hoặc 02 phó giáo sư và 03 tiến sĩ khoa học. Và một số điều kiện khác tại điều 3 khoản 2 các mục a,b,c,d,đ,e của thông tư số 09/2017 bộ GDĐT  |
| 3 | trường hợp nào thì bị đình chỉ tuyển sinh                   | cơ sở đào tạo bị đình chỉ tuyển sinh ngành đào tạo thạc sĩ, tiến sĩ khi để xảy ra một trong các trường hợp sau đây: không đảm bảo một trong các điều kiện mở ngành thạc sĩ, tiến sĩ. Tổ chức tuyển sinh và đào tạo ngoài địa điểm được phép đào tạo. Không đạt điều kiện tiêu chuẩn kiểm định chương trình đào tạo của ngành đăng ký theo quy định của bộ GDĐT, vi phạm pháp luật về giáo dục. Không thực hiện đúng quy định chuyển giao và vi phạm các trường hợp khác theo quy định của pháp luật |
| 4 | thời hạn bị đình chỉ là bao lâu                             | Thời gian đình chỉ tuyển sinh tối thiểu là 12 tháng và tối đa là 24 tháng   |
| 5 | làm cách nào để được tiếp tục tuyển sinh                    | Sau thời hạn đình chỉ tuyển sinh, nếu nguyên nhân dẫn đến việc đình chỉ tuyển sinh được khắc phục và vẫn đảm bảo các điều kiện mở ngành thạc sĩ hoặc tiến sĩ thì bộ trưởng bộ GDĐT quyết định cho phép cơ sở đào tạo được tuyển sinh trở lại  |
| 6 | trường hợp nào thì bị thu hồi quyết định mở ngành           | Bị thu hồi quyết định mở ngành khi xảy ra một trong các trường hợp sau đây: có hành vi gian lận để được mở ngành; vi phạm nghiêm trọng về việc tuyển sinh, quản lý, tổ chức đào tạo; hết thời hạn đình chỉ tuyển sinh mà không khắc phục được nguyên nhân dẫn đến đình chỉ; vi phạm quy định của pháp luật về giáo dục bị xử phạt hành chính ở mức độ phải thu hồi quyết định; các trường hợp khác theo quy định của pháp luật  |

|    |   |  |
|----|---|--|
| 7  | trình tự mở ngành thì như thế nào                             | hội đồng trường(đối với trường công lập), hội đồng quản trị(đối với trường ngoài công lập) có quyết định về mở ngành -> Tổ chức kiểm tra và xác nhận các điều kiện mở ngành -> Thẩm định chương trình đào tạo thạc sĩ hoặc tiến sĩ theo quy định về thẩm định mở ngành -> Gửi hồ sơ mở ngành đào tạo đến cơ quan có thẩm quyền quyết định                          |
| 8  | hồ sơ mở ngành bao gồm những gì                               | hồ sơ bao gồm: công văn đề nghị mở ngành đào tạo, đề án mở ngành, phiếu tự đánh giá điều kiện mở ngành, các tài liệu về xây dựng và thẩm định chương trình đào tạo. Hồ sơ mở ngành được lập thành 2 bản gốc, gửi đến cơ quan có thẩm quyền và phải công bố công khai trên trang thông tin điện tử của cơ sở đào tạo chậm nhất 20 ngày trước khi gửi hồ sơ mở ngành |
| 9  | đề án mở ngành có nội dung như thế nào                        | đề án được xây dựng tại phụ lục I của thông tư số 09/2017/TT-BDGDĐT bao gồm các nội dung: sự cần thiết mở ngành đào tạo, năng lực của cơ sở đào tạo, cơ sở vật chất, thiết bị, chương trình đào tạo, kế hoạch đào tạo và kế hoạch đảm bảo chất lượng đào tạo   |
| 10 | thẩm định chương trình đào tạo là gì                          | sau khi xác nhận các điều kiện đảm bảo chất lượng thực tế, cơ sở đào tạo thành lập hội đồng thẩm định. Quyết định thành lập hội đồng thẩm định phải nêu rõ chức danh, trình độ đào tạo, ngành đào tạo, đơn vị công tác của từng thành viên hội đồng và các điều kiện thẩm định quy định tại thông tư 09/2017/TT-BDGDĐT điều 6, mục 3 và 4                          |
| 11 | tiêu chuẩn của thẩm định chương trình đào tạo thì như thế nào | ngoài tiêu chuẩn thẩm định quy định tại thông tư số 07/2015/TT-BDGDĐT, thành viên hội đồng thẩm định phải là số lẻ, các thành viên còn lại của hội đồng phải đến từ ít nhất 02 cơ sở đào tạo khác, có kinh nghiệm đào tạo cùng ngành hoặc gần ngành và cùng trình độ với chương trình đào tạo được thẩm định. có ít nhất 1 giáo sư hoặc phó giáo sư                |

|    |  |   |
|----|--|---|
| 12 | trong bao lâu thì có quyết định mở ngành | việc xem xét hồ sơ mở ngành đào tạo trình độ thạc sĩ, tiến sĩ được thực hiện trong thời hạn 30 ngày, kể từ ngày nhận được đủ hồ sơ theo quy định của sở đào tạo |
|----|--|---|

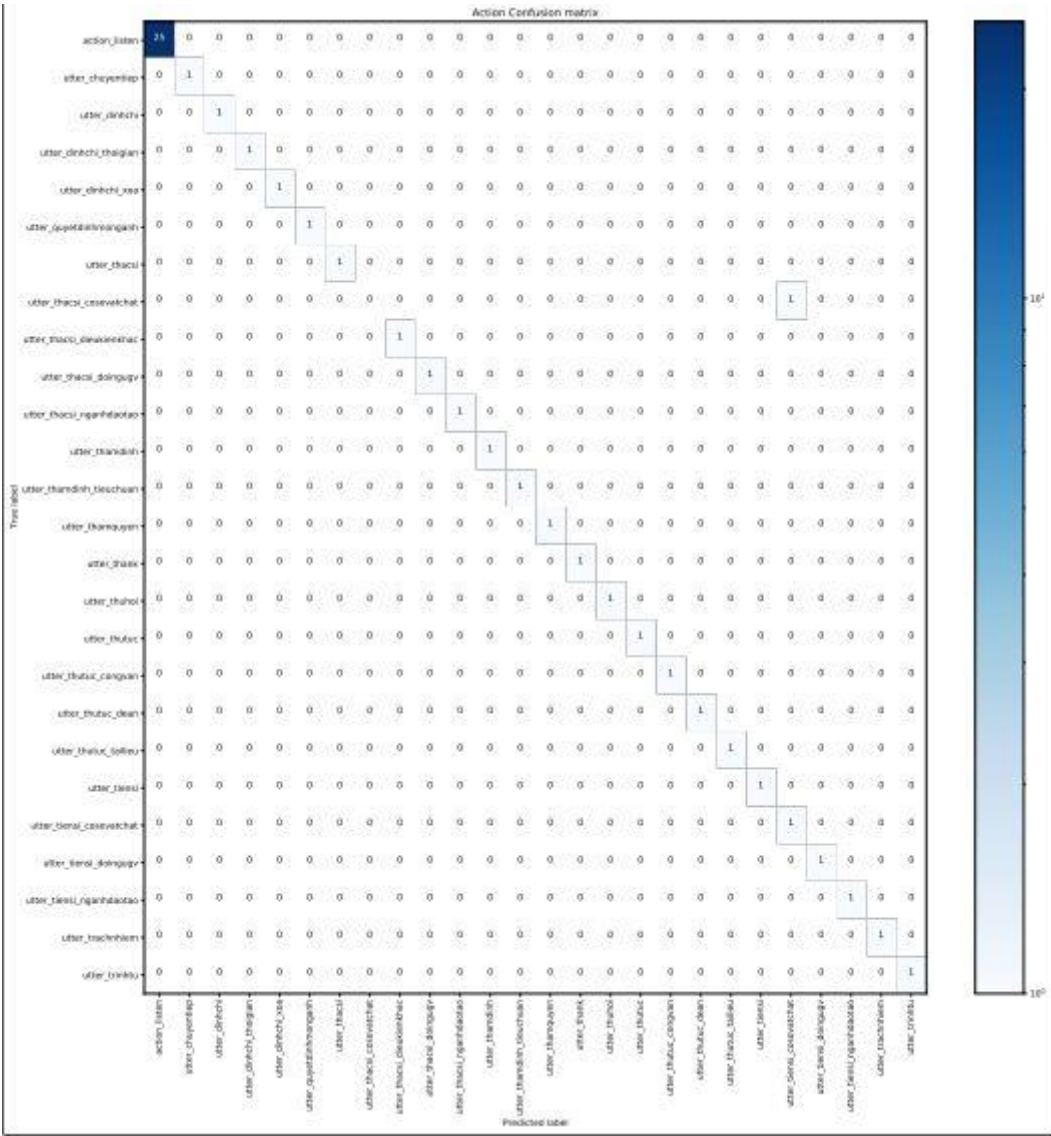
### 3.2 Đánh giá

Một số biểu đồ đánh giá ước lượng tập dữ liệu training và khung kịch bản (history) cho bot:



Hình 3.2: Biểu đồ confusion matrix dữ liệu ngôn ngữ tự nhiên

Biểu đồ confusion matrix của dữ liệu ngôn ngữ tự nhiên cho thấy nhận của các intent thực tế và dự đoán là trùng khớp với nhau. Điều này đảm bảo cho việc nhận dạng chính xác câu hỏi của người dùng (hình 3.2).



Hình 3.3: Confusion matrix dữ liệu các câu trả lời của chatbot

Ứng với mỗi một ý định sẽ có một câu trả lời tương ứng, các câu hỏi có thể khác nhau nhưng cùng gán nhãn ý định thì chỉ có một câu trả lời, với những câu hỏi không có trong phân ý đã dựng sẵn chatbot sẽ điều hướng người sử dụng đến những câu hỏi có sẵn (hình 3.3).

## **PHẦN KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN**

### **1. Kết luận**

Nghiên cứu được thực hiện với mục đích xây dựng chatbot có thể trả lời các câu hỏi pháp luật và tương tác với người sử dụng bằng giọng nói. Qua các kết quả trong quá trình huấn luyện chatbot đề tài đã xây dựng được tập dữ liệu câu hỏi ngôn ngữ tự nhiên phục vụ quá trình huấn luyện chatbot với 163 câu hỏi mẫu và 25 câu trả lời tương ứng trong phạm vi “Thông tư số 09/2017/TT-BGDĐT”. Qua 3 lần kiểm thử với 12 câu hỏi ngẫu nhiên nằm trong kịch bản thu được kết quả 11/12, độ chính xác của bot hiểu được câu hỏi là 91,6% và độ chính xác khi bot phát sinh câu trả lời là 97,3% và có thể tương tác với người sử dụng bằng giọng nói.

### **2. Hướng phát triển**

Quá trình giao tiếp giữa botchat và người sử dụng trải qua rất nhiều giai đoạn dẫn đến bot trả lời rất lâu. Cần cải thiện tốc độ phản hồi của bot.

Vì giới hạn của đề tài nên việc giao tiếp với bot chỉ có thể thực hiện trên máy tính cá nhân và phải cài đầy đủ các thư viện cần có. Do đó cần triển khai chatbot lên một nền tảng trực tuyến để người sử dụng có thể dễ dàng tương tác.

Tại một thời điểm bot chỉ có thể trả lời một câu hỏi duy nhất, nên cần xây dựng hệ thống multi intent để bot có thể trả lời được những câu hỏi ghép.



## TÀI LIỆU THAM KHẢO

### Tài liệu trong nước

Đỗ Thanh Nghị, Hoàng Tùng. *Chatbot cho sinh viên công nghệ thông tin*, kỷ yếu Hội nghị Quốc gia lần thứ XII về nghiêm cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR) Huế, 2019.

Thông tư số: 09/2017/TT-BGDĐT ngày 4 tháng 4 năm 2017. *Về việc quy định điều kiện, trình tự, thủ tục mở ngành hoặc chuyên ngành đào tạo và đình chỉ tuyển sinh, thủ tục hỏi quyết định mở ngành hoặc chuyển ngành đào tạo trình độ thạc sĩ, trình độ tiến sĩ.*

### Tài liệu ngoài nước

Anbang Xu, Anbang Xu Zhe, Liu, Zhe Liu, Yufan Guo, Yufan Guo, Vibha Singhal Sinha, Vibha Sinha, Rama Akkiraju, Rama Akkiraju. *A New Chatbot for Customer Service on Social Media*, May 2017.

Ledell Wu, Adam Fisch, Sumit Chopra, Keith Adams, Antoine Bordes and Jason Weston. *StarSpace: Embed All The Things!*, 21 Nov 2017.