**TÀI LIỆU CHATBOT HỖ TRỢ ĐẢNG VIÊN**

I/ Yêu cầu bài toán

Xây dựng một chatbot trả lời câu hỏi của đảng viên.

II/ Luồng hoạt động

Diagram

Description automatically generated

*Hình 1. Luồng hoạt động của chatbot hỗ trợ Đảng viên*

III/ Công nghệ được sử dụng

1. FastText

FastText là một thư viện do Facebook phát triển đạt hiệu quả trong việc biểu diễn từ và cả phân loại câu theo chủ đề.

Để tìm hiểu thêm thông tin vui lòng tham khảo link tại [đây](https://fasttext.cc/).

Ở dự án này, chúng tôi sử dụng FastText với mục đích phân loại câu hỏi theo các chủ đề (Sentences Classification) nhằm thuận tiện cho việc tìm kiếm trong cơ sở dữ liệu kịch bản.

1. BERT

BERT là model biểu diễn ngôn ngữ được Google giới thiệu vào năm 2018.

Để tìm hiểu thêm thông tin vui lòng tham khảo link tại [đây](https://arxiv.org/abs/1810.04805).

Ở dự án này, chúng tôi sử dụng một biến thể của BERT với khả năng chuyển đổi câu, model này chuyển câu thành một vector để phục vụ cho việc tìm câu đồng nghĩa. Model được chúng tôi sử dụng là [model này](https://huggingface.co/keepitreal/vietnamese-sbert).

IV/ Mô tả sản phẩm

- Một bộ API gồm:

+ API trả lời câu hỏi của đảng viên.

+ API CRUD dữ liệu (cụ thể là kịch bản trả lời câu hỏi) trong Database.

V/ Cách cài đặt

1. Cài đặt môi trường và các thư viện cần thiết.

1.1. Tạo môi trường ảo Venv

a. Windows

python -m venv venv

venv/Scripts/activate

b. Linux (Ubuntu)

python -m venv \venv

source venv/bin/activate

1.2. Cài đặt thư viện cần thiết

pip install -r requirements.txt

**# Lưu ý:**

- Do module FastText có thể gặp lỗi khi sử dụng phương pháp install **pip** hay **pip3** của Python, nên ở đây chúng tôi khuyến khích bạn đọc tài liệu hướng dẫn install của chính FastText tại [đây](https://fasttext.cc/docs/en/supervised-tutorial.html).

1.3. Huẩn luyện cho mô hình FastText.

Trong module thêm tệp dữ liệu cho việc huấn luyện, có tên là “train\_data\_final.txt”. Về format của dữ liệu, bạn có thể tham khảo chính file “train\_data\_final.txt” của chúng tôi. Sau đó bạn chạy file “model\_fasttext.py”, hệ thống sẽ tự sinh một model “model\_dangvien.bin”, hoặc bạn có thể sử dụng luôn model chúng tôi đã train cùng tên.

2. Cách chạy hệ thống

python main.py

# Setup Database.

- Ở đây sử dụng MongoDB.

- Với tập dữ liệu chứa trong Collection ở local.

- Một document gồm một kịch bản trả lời câu hỏi:

+ '\_id' : id của kịch bản.

+ 'label' : Nhãn của câu hỏi.

+ 'question' : Câu hỏi.

+ 'answer' : Câu trả lời.

+ 'link' : Đường link tài liệu tham khảo.

+ 'vector' : Một mảng (Array) được tạo ra từ việc nhúng (Embedding) câu hỏi.