**BÁO CÁO TIẾN ĐỘ THỰC HIỆN ĐỒ ÁN TỐT NGHIỆP**

**ĐỀ TÀI: XÂY DỰNG WEBSITE**

**HỎI ĐÁP TRỰC TUYẾN VỀ CÔNG NGHỆ**

------------------------

**GVHD:** TS. Huỳnh Xuân Phụng

**SVTH:** Trấn Tấn Quý

Phạm Hoàng Việt

1. **PHẠM VI NGHIÊN CỨU**
2. **MỤC TIÊU**

Xây dựng website hỏi đáp trực tuyến về công nghệ.

1. **NỘI DUNG**

- Xây dựng chức năng hỏi đáp, bảng xếp hạng câu hỏi.

- Xây dựng chức năng quản lý người dùng, câu hỏi.

- Tìm hiểu về recommendation system và xử lý ngôn ngữ tự nhiên: mô hình Word2Vec.

- Training dữ liệu câu hỏi tiếng Anh từ Google Big Query Dataset.

- Xây dựng API:

+ Tìm hiểu Flask Framework trong Python.

+ Xây dựng API cho việc tương tác gợi ý câu hỏi

+ Kết hợp với MongoDB để thực hiện việc training dữ liệu khi phát sinh dữ liệu mới.

- Thực hiện chức năng gợi ý câu hỏi trên Laravel .

- Xây dựng chức năng chat real time.

- Tính năng tìm kiếm nâng cao.

1. **KHỐI LƯỢNG CÔNG VIỆC ĐÃ THỰC HIỆN**

- Tìm hiểu kiến thức liên quan:

+ Natural Language Processing (NLP): xử lý ngôn ngữ tự nhiên.

+ Word2Vec: mô hình hóa từ ngữ thành không gian vector.

+ Long Short Term Memory Model.

- Thu thập dataset từ Google BigQuery

- Tiền xử lý dữ liệu.

- Training tập dữ liệu

- Xây dựng API sử dụng Flask Framework và test API sử dụng Postman

- Tìm hiểu và deploy được project lên docker.

**3. KHÓ KHĂN**

- Việc training đòi hỏi máy tính cần cấu hình có GPU và cài đặt một số thư viện hỗ trợ việc training dữ liệu.

- Khi có dữ liệu mới được thêm vào hệ thống, việc training dữ liệu cần được thực hiện lại. Tuy nhiên, nhóm vẫn chưa tìm được cách để thực hiện việc training hiệu quả nhằm đảm bảo dữ liệu (đã qua training) được đưa lên hệ thống một cách đầy đủ và nhanh nhất có thể. Việc training ngay lập tức khi dữ liệu mới được đưa vào hệ thống không khả thi vì mỗi lần training sẽ mất một khoảng thời gian tương đối dài.

*TP. Hồ Chí Minh, ngày 01 tháng 05 năm 2020*

**GIẢNG VIÊN HƯỚNG DẪN SINH VIÊN THỰC HIỆN**

TS. Huỳnh Xuân Phụng Trần Tấn Quý Phạm Hoàng Việt