BÁO CÁO TIẾN ĐỘ TỔNG KẾT 3 THÁNG THỰC TẬP

Người làm: Trần Đức Thọ

-Những thứ em đã tìm hiểu được trong 3 tháng thực tập vừa qua:

+ Ôn lại cấu trúc dữ liệu và thuật toán.

+ Giải quyết được bài toán tìm kiếm các từ trong wordDict trong chuỗi, tìm các từ trong ma trận các chữ. . .

+ Tìm hiểu về visualize data bằng thư viện matplotlib, hiểu cách biểu diễn văn bản bằng n-gram, tf-idf và demo

+ Tìm hiểu được thuật toán SVM

+ Tìm hiểu được giải thuật LightGBM áp dụng LightGBM giải quyết bài toán “Click-Through Rate Prediction”

+ Hiểu về L1 và L2 regularization

+ Tìm hiểu được về phương pháp giảm chiều dữ liệu PCA.

+ Tìm hiểu về mô hình deep learning: mạng RNN, mạng LSTM, mạng transformer

+ Tìm hiểu mã nguồn mở OpenNMT: cách cài đặt file config, cách train mô hình và cách dùng mô hình train được để dịch văn bản từ tiếng anh sang tiếng việt . . .

+ Tạo được 1 API dịch từ tiếng anh sang tiếng việt:

Graphical user interface, application

Description automatically generated

Việc còn chưa hoàn thành:

+ Chưa tìm hiểu về Option về công nghệ + design system.

Cách chạy api:

Clone code từ trang github: <https://github.com/weedharm/apien-vi>

Tải model từ địa chỉ :

Sau khi hoàn thành sẽ có giao diện như sau

A screenshot of a computer

Description automatically generated

Chạy terminal:

Streamlit run dich.py

Sẽ hiển thị giao diện tại địa chỉ: <http://localhost:8501/>

Graphical user interface, application

Description automatically generated

Nhập câu tiếng anh cần tìm kiếm và ấn translate:

Graphical user interface, application

Description automatically generated

Máy sẽ trả về câu cần dịch:

Graphical user interface, application

Description automatically generated

Cách 2: chỉ cần clone về và chạy terminal

Streamlit run test.py

Cũng sẽ trả ngay ra giao diện như trên.