

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA KHOA HỌC VÀ KỸ THUẬT MÁY TÍNH



Thực tập ngoài trường - CO3335

Báo cáo

AUTOFILL SERVICE

Giảng viên hướng dẫn: Trương Quỳnh Chi

Sinh viên thực hiện: 2111762 - Phạm Võ Quang Minh

TP. Hồ Chí Minh, 12/2024

Mục lục

1 Giới thiệu doanh nghiệp	3
2 Nội dung thực tập	4
2.1 Giới thiệu đề tài	4
2.2 Phân tích chi tiết đề tài	4
2.2.1 Biến dữ liệu chưa có cấu trúc về dữ liệu có cấu trúc	4
2.2.2 Dịch vụ API	4
2.3 Hướng tiếp cận và giải quyết	4
2.3.1 Biến dữ liệu chưa có cấu trúc về dữ liệu có cấu trúc	4
2.3.1.a Tiền xử lí	4
2.4 Hiện thực	4
3 Kết quả đạt được	5
4 Kết luận	6

Danh mục hình ảnh

Hình 1: Logo doanh nghiệp Apollogix	3
---	---

Danh mục bảng biểu

1 Giới thiệu doanh nghiệp

Công ty TNHH Apollogix chuyên cung cấp các giải pháp phần mềm quản lý chuyên dụng trong lĩnh vực giao thông vận tải Thông tin công ty:

- Tên công ty: Công ty TNHH Apollogix
- Tên quốc tế: APOLLOGIX COMPANY LIMITED
- Tên viết tắt: APOLLOGIX CO LTD
- Trụ sở chính: 39 Đường B4, phường An Lợi Đông, Quận 2, thành phố Thủ Đức.
- Văn phòng làm việc: 39 Đường B4, phường An Lợi Đông, Quận 2, thành phố Thủ Đức.
- Điện thoại: 0796513044 (Ms. Huyền).
- Email: contact@apollogix.com.
- Weblink: apollogix.com.
- Logo nhận diện:



Hình 1: Logo doanh nghiệp Apollogix

2 Nội dung thực tập

2.1 Giới thiệu đề tài

Trong hệ thống TMS có một chức năng gọi là Autofill, chức năng này có input đầu vào là file pdf booking, output mong đợi là đọc từ file pdf này ra cấu trúc json của dữ liệu Transport Job đang có để sau đó sẽ hiển thị data này trên giao diện form Transport Job cho end user xem trước khi lưu lại mà không cần phải nhập. Ngoài ra dịch vụ có thể được sử dụng để tạo đơn hàng có trạng thái chờ xác nhận từ email của khách hàng hoặc trên customer portal sau này.

File PDF Booking này khách hàng có được từ bên các hãng tàu hoặc các dịch vụ bên thứ 3 cung cấp, có chứa các thông tin về lịch cảng tàu, thông tin container cần vận chuyển, v.v

Sinh viên thực hiện: Phạm Võ Quang Minh

Nhân sự hỗ trợ kỹ thuật: Anh Thi

Nhân sự hỗ trợ nghiệp vụ: Anh Phúc, Chị Nguyệt, và các nhân sự bên vận hành

2.2 Phân tích chi tiết đề tài

Đề tài có thể được chia ra làm hai bước chính, mỗi bước sẽ bao gồm hai mảng kiến thức khác nhau.

2.2.1 Biến dữ liệu chưa có cấu trúc về dữ liệu có cấu trúc

Phát triển một hàm có khả năng biến dữ liệu từ định dạng pdf (dữ liệu không cấu trúc) sang dạng object (dữ liệu có cấu trúc).

2.2.2 Dịch vụ API

Phát triển dịch vụ (API service) để nhận đầu vào từ nơi gọi (hệ thống TMS) và trả về dữ liệu có cấu trúc cho nơi gọi.

2.3 Hướng tiếp cận và giải quyết

2.3.1 Biến dữ liệu chưa có cấu trúc về dữ liệu có cấu trúc

Đề xuất ba phương pháp có thể sử dụng:

1. Phương pháp xử lý truyền thống bằng lý thuyết xử lý ngôn ngữ tự nhiên
2. Sử dụng mô hình ngôn ngữ lớn (LLM) như ChatGPT để trích xuất thông tin. Trong đó có yêu cầu mô hình ngôn ngữ trả về dạng chuỗi (string) theo format của của 1 dạng dữ liệu có cấu trúc như json.

2.3.1.a Tiềm xử lý

2.4 Hiện thực



3 Kết quả đạt được



4 Kết luận





TÀI LIỆU THAM KHẢO