TMA Solutions

FINAL REPORT

NGUYEN DANG MAI THY

AI Center, TMA Innovations

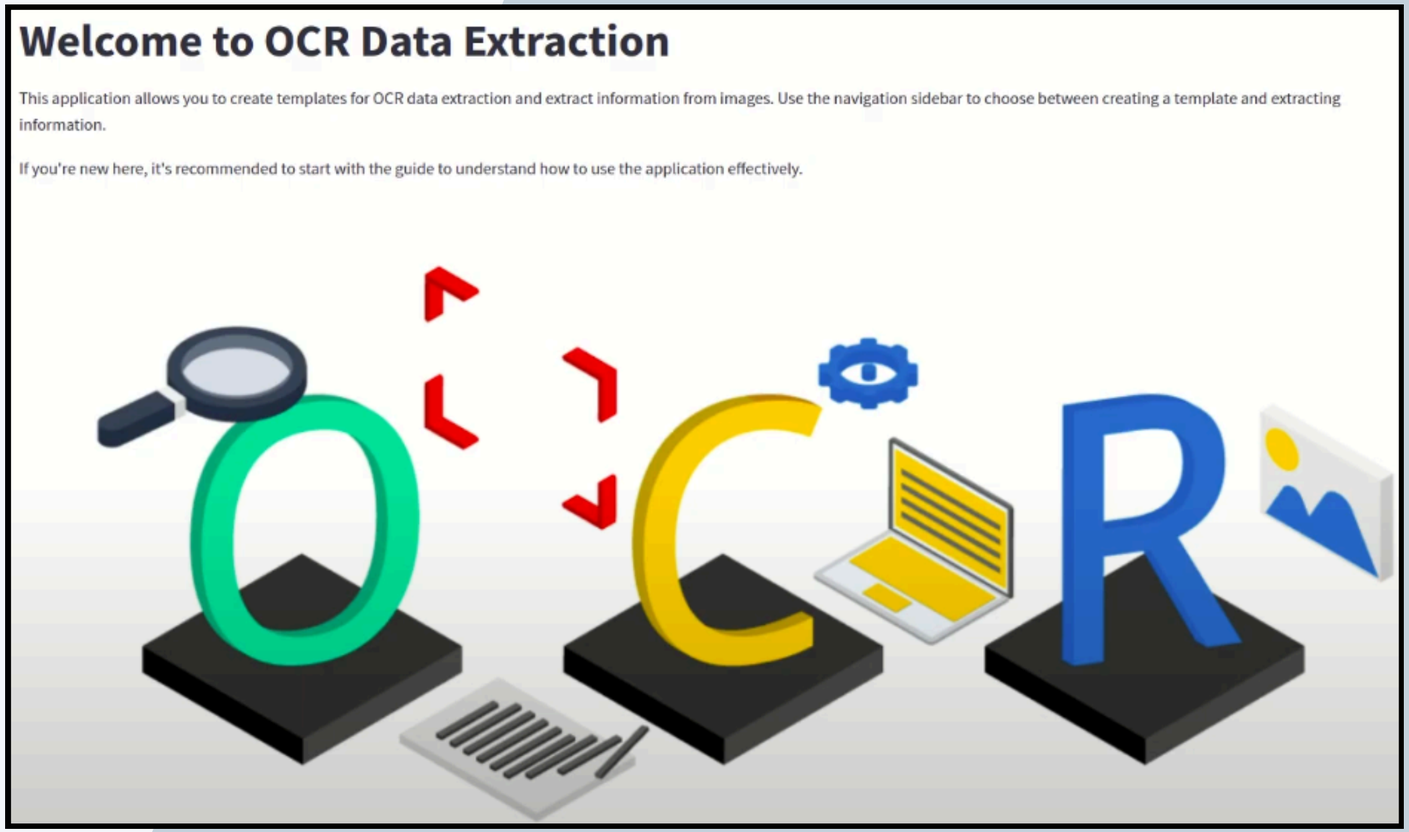
TMA Solutions

Introduction

I am Nguyen Dang Mai Thy, an intern from Batch 43 of the TMA Internship Program.

I was assigned to the AI Center, working in the Computer Vision domain, with Tran Trung Kien as my mentor and Nguyen Minh Nghia as the Project Manager.

I was allocated to the Dynamic Template OCR project and also provided support for several related projects.



Dynamic Template OCR is a unique software that allows users to create custom OCR templates to process complex and unstructured documents.

Users can easily design and adjust the OCR templates to fit unique document types, such as contracts, reports, and invoices.

The custom OCR templates can identify and extract key data fields from documents, even when they lack clear structure.

TMA Solutions

Requirements of Dynamic Template OCR

Project Scope:

- The project covers processing of input images and PDF documents in Vietnamese and English.

Project Requirements:

- Image preprocessing
- Document type recognition to classify into pre-existing templates
- Template creation including key-value pair extraction
- Table detection and table reconstruction
- Information extraction - Text OCR

TMA Solutions

Key deliverables

- Implement template grouping for template creation module
- Implement template grouping for information extraction
- Prepare data for table reconstruction
- Table Detection & Table Classification
- Table Reconstruction
- Integrate the project with Database and API
- Integrate the project with a user interface

TMA Solutions

Result Achievements

- Research on Table Recognition (TR)
- Test & Compare TR repository
- Collect data to evaluate accuracy for TR model
- Draw Dynamic OCR full-flow
- Modify & Optimize TR code
- Collect data for training Label Medicine Box
- Research on PostgreSQL
- Research on FlaskAPI

TMA Solutions

Research on Table Recognition

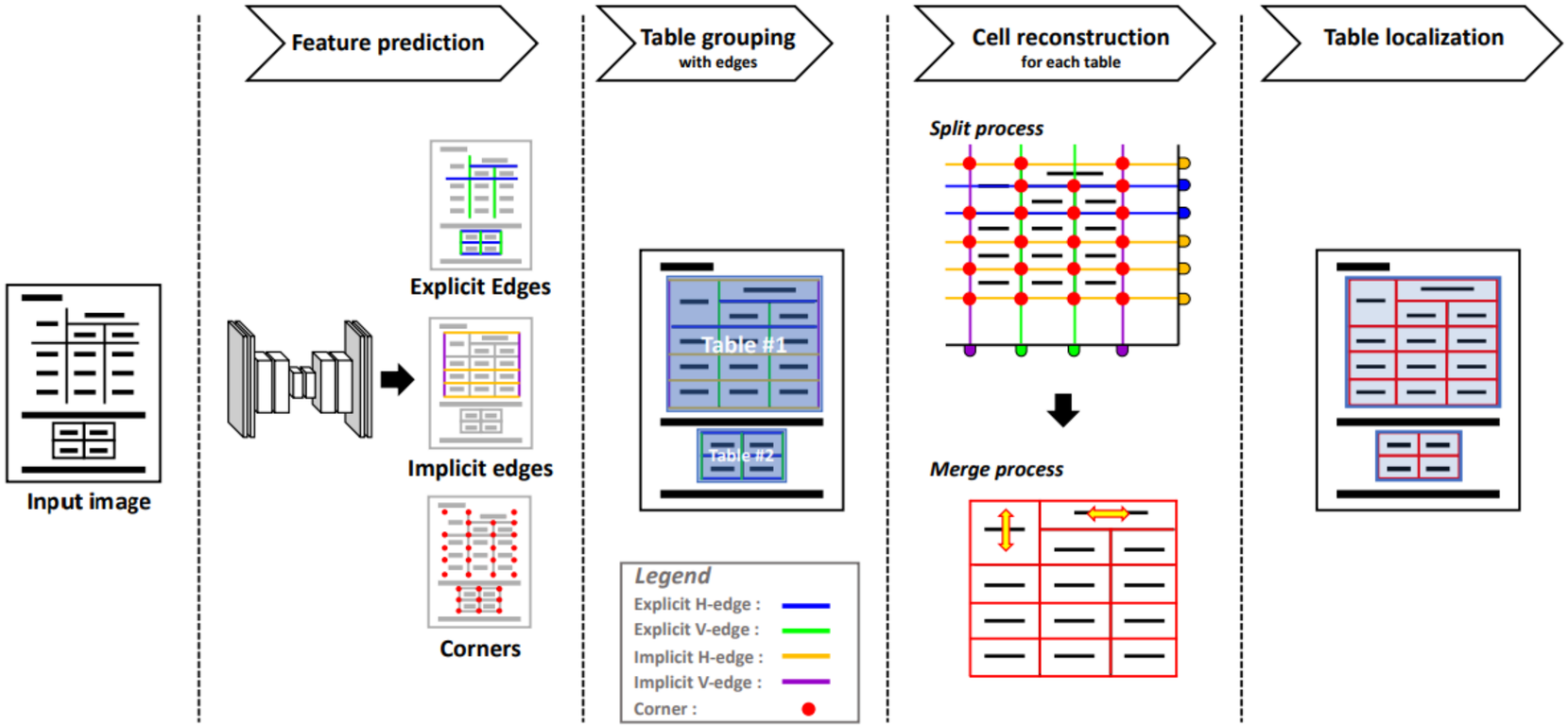


Table Recognition is a challenging problem within the broader field of Document Understanding. The goal is to accurately detect, recognize, and extract information from tabular data in documents, while also preserving the layout and structure of the table.

Key sub-tasks include Table Detection, Table Structure Recognition, Table Recognition, and the most complex, Table Understanding, which may involve extracting information from the table based on additional inputs like queries.

Document Understanding is a broad field that uses AI and automation to extract information from various document types, both structured and unstructured.

TMA Solutions

Test & Compare TR repository

Dataset	TD	TSR	TR	Number of sample	TSR Format
TableBank	✓	✓	✗	417K (TD), 144k (TSR)	Table Detection, Table2HTML
PubTabNet	✗	✓	✓	568K	Table2HTML
FinTabNet	✗	✓	✓	113K	Table2HTML
SciTSR	✗	✓	✓	15K	Text bounding box, Relative cell position, Cell Adjacency Relation
TIES_2.0	✗	✓	✓	Unbounded	Synthetic data, table-type classification, Table Structure to HTML
TabLex	✗	✓	✓	1M+	Structure information code, Content information code (similar to Table2HTML)
PubTable	✓	✓	✓	948K	Structure recognition, functional analysis, text content & bbox location
...					

Dataset	TSR Format	Note
PubTabNet	Table2HTML	1/2 are span-cells + majority are borderless ⇒ when the input is full-border, the accuracy is low (majority of Vietnamese data is full-border)
TableBank	Table Detection, Table2HTML	
SciTSR	Text bounding box, Relative cell position, Cell Adjacency Relation	
FinbTabNet	Table2HTML	
TNCR	5 class	

For the pipeline, the Table Detection model can be trained on a composite dataset combining resources like TableBank, FinTabNet, PubLayNet, and TNCR.
The model will detect a single table class, with a separate model used to classify the table type in a subsequent step.

TMA Solutions

Test & Compare TR repository

Repository	Advantages	Disadvantages
PaddleOCR-ppstructer	- Lightweight and optimized model - Supports direction correction	- Limited dataset - Difficult to develop and expand - Challenging post-processing - Poor Vietnamese language support
TableMaster, MTL TabNet, EDD-third-party, tsr-convstem	N/A	- Difficult to create dataset - Heavy models - Extremely difficult and infeasible to create more data
GraphTSR	N/A	- Difficult to prepare the data - Unrealistic assumptions about the input
deepdoctection	- Can be developed and expanded the best as it allows customization of the format and use of custom data	- Default uses the PubTabNet dataset, so it cannot detect borderless tables, but can be retrained with custom data to improve - Works well on the demo app, good Vietnamese language support
Davar-OCR	N/A	- Quite complex design with many sub-modules
DeepTSR, TableNet	N/A	- The model is too simple, unable to detect span-cells

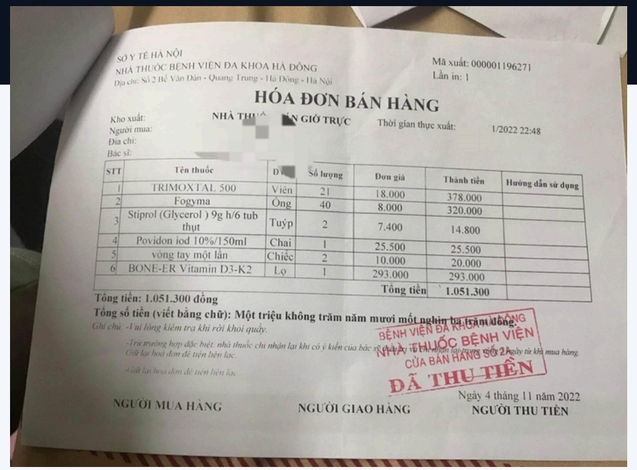
The comparison indicates that while DeepDocDetection has strong table detection and cell detection, its table reconstruction is weaker and it cannot handle span cells.

After further research,TRACE appears to be the best performing in terms of detecting and redrawing span cells, as well as overall table reconstruction, outperforming all the other remaining repositories.

TMA Solutions

Collect data to evaluate accuracy for TR model

- TB-BangDiem-8
- TB-Donthuoc-9
- TB-GiayDkyThue-6
- TB-GiayHosoXeOto-3
- TB-Giaykhamsuckhoe-5
- TB-Giaynoptien-7
- TB-GiayThaydoiKinhdoanh-3
- TB-Hopdongmuaban



SỞ Y TẾ HÀ NỘI
NHÀ THUỐC BỆNH VIỆN ĐA KHOA HÀ ĐÔNG
Địa chỉ: Số 2 Hồ Văn Dãn - Quang Trung - Hà Đông - Hà Nội

Mã xuất: 000001196271
Lần in: 1

HÓA ĐƠN BÁN HÀNG

NHÀ THUỐC BỆNH VIỆN ĐA KHOA HÀ ĐÔNG TRỰC
Thời gian thực xuất: 1/2022 22:48

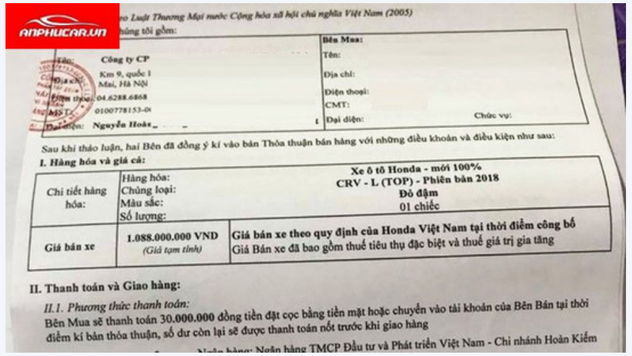
Kho xuất:
Người mua:
Địa chỉ:
Bác sĩ:

STT	Tên thuốc	Đ	Số lượng	Đơn giá	Thành tiền	Hướng dẫn sử dụng
1	TRIMOXTAL 500	Viên	21	18.000	378.000	
2	Fogyma	Ống	40	8.000	320.000	
3	Stiprol (Glycerol) 9g h/6 tub thụt	Tuýp	2	7.400	14.800	
4	Povidon iod 10%/150ml	Chai	1	25.500	25.500	
5	vòng tay một lần	Chiếc	2	10.000	20.000	
6	BONE-ER Vitamin D3-K2	Lọ	1	293.000	293.000	
				Tổng tiền	1.051.300	

Tổng tiền: 1.051.300 đồng
Tổng số tiền (viết bằng chữ): Một triệu không trăm năm mươi một nghìn ba trăm đồng.
Ghi chú: -Vui lòng kiểm tra khi rời khỏi quầy.
-Trừ trường hợp đặc biệt, nhà thuốc chỉ nhận lại khi có ý kiến của bác sĩ điều trị. Quý khách vui lòng giữ lại hóa đơn để nhận lại tiền.
-Giữ lại hóa đơn để tiện liên lạc.

NHÀ THUỐC BỆNH VIỆN ĐA KHOA HÀ ĐÔNG
CỬA BÁN HÀNG SỐ 2
ĐÃ THU TIỀN
Ngày 4 tháng 11 năm 2022

NGƯỜI MUA HÀNG | NGƯỜI GIAO HÀNG | NGƯỜI THU TIỀN



ANPHUCAN.VN

...o Luật Thương Mại nước Cộng hòa xã hội chủ nghĩa Việt Nam (2005)
...húng tôi gồm:

Bên Bán:
Tên: Công ty CP
Địa chỉ: Km 9, quốc l Mai, Hà Nội
Điện thoại: 04.6288.6868
CMT: 0100778153-06
Đại diện: Nguyễn Hoàn... Chức vụ:

Bên Mua:
Tên:
Địa chỉ:
Điện thoại:
CMT:
Đại diện:

Sau khi thảo luận, hai Bên đã đồng ý kí vào bản Thỏa thuận bán hàng với những điều khoản và điều kiện như sau:

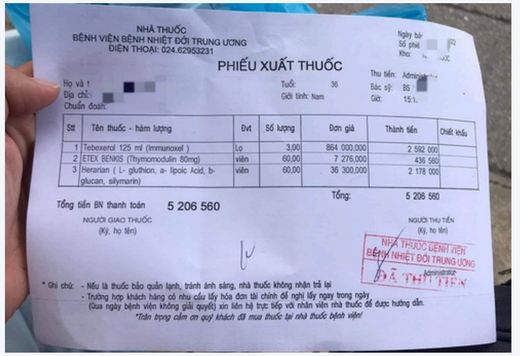
I. Hàng hóa và giá cả:

Chi tiết hàng hóa:	Hàng hóa:	Xe ô tô Honda - mới 100%
	Chủng loại:	CRV - L (TOP) - Phiên bản 2018
	Màu sắc:	Đỏ đậm
	Số lượng:	01 chiếc
Giá bán xe	1.088.000.000 VND (Giá tạm tính)	Giá bán xe theo quy định của Honda Việt Nam tại thời điểm công bố. Giá Bán xe đã bao gồm thuế tiêu thụ đặc biệt và thuế giá trị gia tăng

II. Thanh toán và Giao hàng:

II.1. Phương thức thanh toán:
Bên Mua sẽ thanh toán 30.000.000 đồng tiền đặt cọc bằng tiền mặt hoặc chuyển vào tài khoản của Bên Bán tại thời điểm kí bản thỏa thuận, số dư còn lại sẽ được thanh toán nốt trước khi giao hàng

...Ngân hàng TMCP Đầu tư và Phát triển Việt Nam - Chi nhánh Hoàn Kiếm



NHÀ THUỐC
BỆNH VIỆN BỆNH NHIỆT ĐỚI TRUNG ƯƠNG
ĐIỆN THOẠI: 024.62953231

Ngày bán:
Số phiếu:
Kho:

PHIẾU XUẤT THUỐC

Họ và t...
Địa chỉ:
Chuẩn đoán:

Thu tiền: Administrator
Tuổi: 36 Bác sỹ: BS
Giới tính: Nam Giờ: 15:1...

Stt	Tên thuốc - hàm lượng	Đvt	Số lượng	Đơn giá	Thành tiền	Chiết khấu
1	Tebexerol 125 ml (Immunoxel)	Lọ	3,00	864 000,000	2 592 000	
2	ETEX BEMIS (Thymomodulin 80mg)	viên	60,00	7 276,000	436 560	
3	Heranran (L- glutihion, a- lipoic Acid, b-glucan, silymarin)	viên	60,00	36 300,000	2 178 000	
				Tổng:	5 206 560	

Tổng tiền BN thanh toán 5 206 560

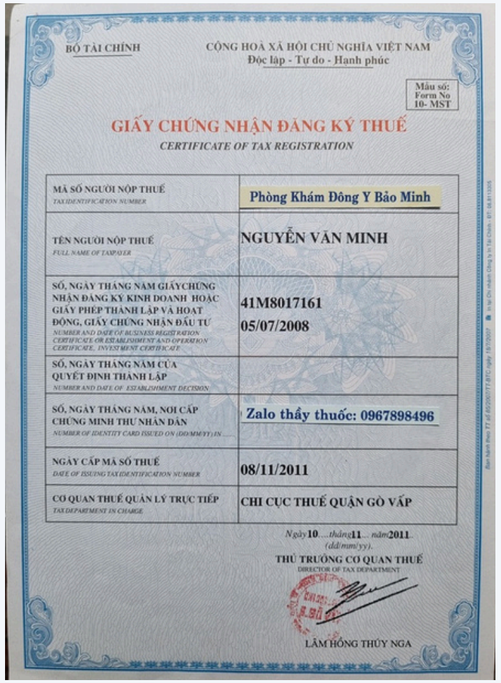
NGƯỜI GIAO THUỐC
(Ký, họ tên)

NGƯỜI THU TIỀN
(Ký, họ tên)

NHÀ THUỐC BỆNH VIỆN BỆNH NHIỆT ĐỚI TRUNG ƯƠNG
Administrator
ĐÃ THU TIỀN

* Ghi chú: - Nếu là thuốc bảo quản lạnh, tránh ánh sáng, nhà thuốc không nhận trả lại
- Trường hợp khách hàng có nhu cầu lấy hóa đơn tài chính đề nghị lấy ngay trong ngày
(Qua ngày bệnh viện không giải quyết) xin liên hệ trực tiếp với nhân viên nhà thuốc để được hướng dẫn.
*Trân trọng cảm ơn quý khách đã mua thuốc tại nhà thuốc bệnh viện!

TT	Ngành đào tạo	Mã ngành	Mã tổ hợp xét tuyển	Điểm chuẩn	
KHỐI KINH TẾ, KỸ THUẬT, KHOA HỌC XÃ HỘI – NHÂN VĂN					
1.	Kế toán	7340301	A00, A01, D01	15.0	15.5
2.	Tài chính - Ngân hàng	7340201	A00, A01, D01	15.0	15.5
3.	Quản trị kinh doanh	7340101	A00, A01, D01	15.0	15.5
4.	Quản trị văn phòng	7340406	A00, A01, D01, C00	15.0	15.5
5.	Kỹ thuật điện	7520201	A00, A01	15.0	15.5
6.	Công nghệ thông tin	7480201	A00, A01, D01	15.0	15.5
7.	Kinh tế	7310101	A00, A01, D01	15.0	15.5
8.	Ngôn ngữ Anh	7220201	D01, D09, D10, A01	15.0	15.5
9.	Marketing	7340115	A00, A01, D01	15.0	15.5
10.	Công nghệ kỹ thuật điện tử - viễn thông	7510302	A00, A01, D01	15.0	15.5



BỘ TÀI CHÍNH
CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

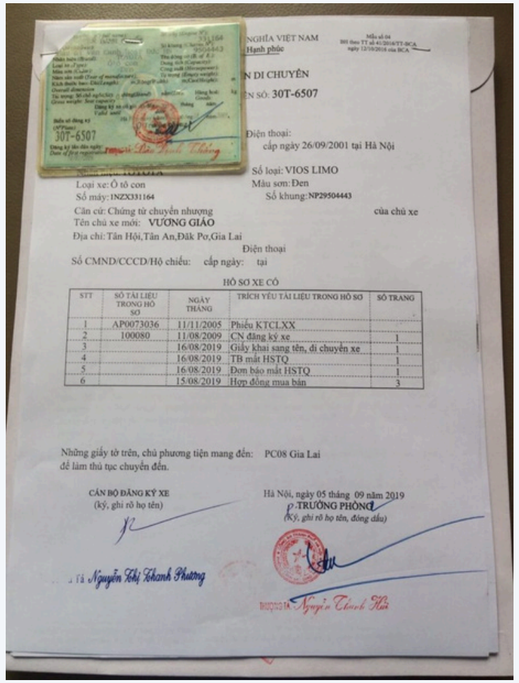
Mẫu số: Form No
10- MST

GIẤY CHỨNG NHẬN ĐĂNG KÝ THUẾ
CERTIFICATE OF TAX REGISTRATION

MÃ SỐ NGƯỜI NỘP THUẾ *TAX IDENTIFICATION NUMBER*	Phòng Khám Đông Y Bảo Minh
TÊN NGƯỜI NỘP THUẾ *FULL NAME OF TAXPAYER*	NGUYỄN VĂN MINH
SỐ, NGÀY THÁNG NĂM GIẤY CHỨNG NHẬN ĐĂNG KÝ KINH DOANH HOẶC GIẤY PHÉP THÀNH LẬP VÀ HOẠT ĐỘNG, GIẤY CHỨNG NHẬN ĐẦU TƯ *NUMBER AND DATE OF BUSINESS REGISTRATION CERTIFICATE OR ESTABLISHMENT AND OPERATION CERTIFICATE, INVESTMENT CERTIFICATE*	41M8017161 05/07/2008
SỐ, NGÀY THÁNG NĂM CỦA QUYẾT ĐỊNH THÀNH LẬP *NUMBER AND DATE OF ESTABLISHMENT DECISION*	
SỐ, NGÀY THÁNG NĂM, NƠI CẤP CHỨNG MINH THƯ NHÂN DÂN *NUMBER OF IDENTITY CARD ISSUED ON (DD/MM/YY) IN...*	Zalo thầy thuốc: 0967898496
NGÀY CẤP MÃ SỐ THUẾ *DATE OF ISSUING TAX IDENTIFICATION NUMBER*	08/11/2011
CƠ QUAN THUẾ QUẢN LÝ TRỰC TIẾP *TAX DEPARTMENT IN CHARGE*	CHI CỤC THUẾ QUẬN GÒ VẤP

Ngày 10 tháng 11 năm 2011
(dd/mm/yy)
THỦ TRƯỞNG CƠ QUAN THUẾ
DIRECTOR OF TAX DEPARTMENT

LÂM HỒNG THÚY NGA



...NGHĨA VIỆT NAM
...Hạnh phúc

...N DI CHUYỂN
...N SỐ: 30T-6507

Điện thoại:
...cấp ngày 26/09/2001 tại Hà Nội

Loại xe: Ô tô con
Số máy: INZX331164
Số loại: VIOS LIMO
Màu sơn: Đen
Số khung: NP29504443

Căn cứ: Chứng từ chuyển nhượng
Tên chủ xe mới: VƯƠNG GIÁO
của chủ xe
Địa chỉ: Tân Hội,Tân An,Đắk Pơ,Gia Lai
Điện thoại
Số CMND/CCCD/Hộ chiếu: cấp ngày: tại

HỒ SƠ XE CÓ

STT	SỐ TÀI LIỆU TRONG HỒ SƠ	NGÀY THÁNG	TRÍCH YẾU TÀI LIỆU TRONG HỒ SƠ	SỐ TRANG
1	AP0073036	11/11/2005	Phiếu KTCLXX	
2	100080	11/08/2009	CN đăng ký xe	1
3		16/08/2019	Giấy khai sang tên, di chuyển xe	1
4		16/08/2019	TB mất HSTQ	1
5		16/08/2019	Đơn báo mất HSTQ	1
6		15/08/2019	Hợp đồng mua bán	3

Những giấy tờ trên, chủ phương tiện mang đến: PC08 Gia Lai
để làm thủ tục chuyển đến.

CÁN BỘ ĐĂNG KÝ XE
(ký, ghi rõ họ tên)

Hà Nội, ngày 05 tháng 09 năm 2019
P.TRƯỞNG PHÒNG
(Ký, ghi rõ họ tên, đóng dấu)

Nguyễn Thị Thanh Phương

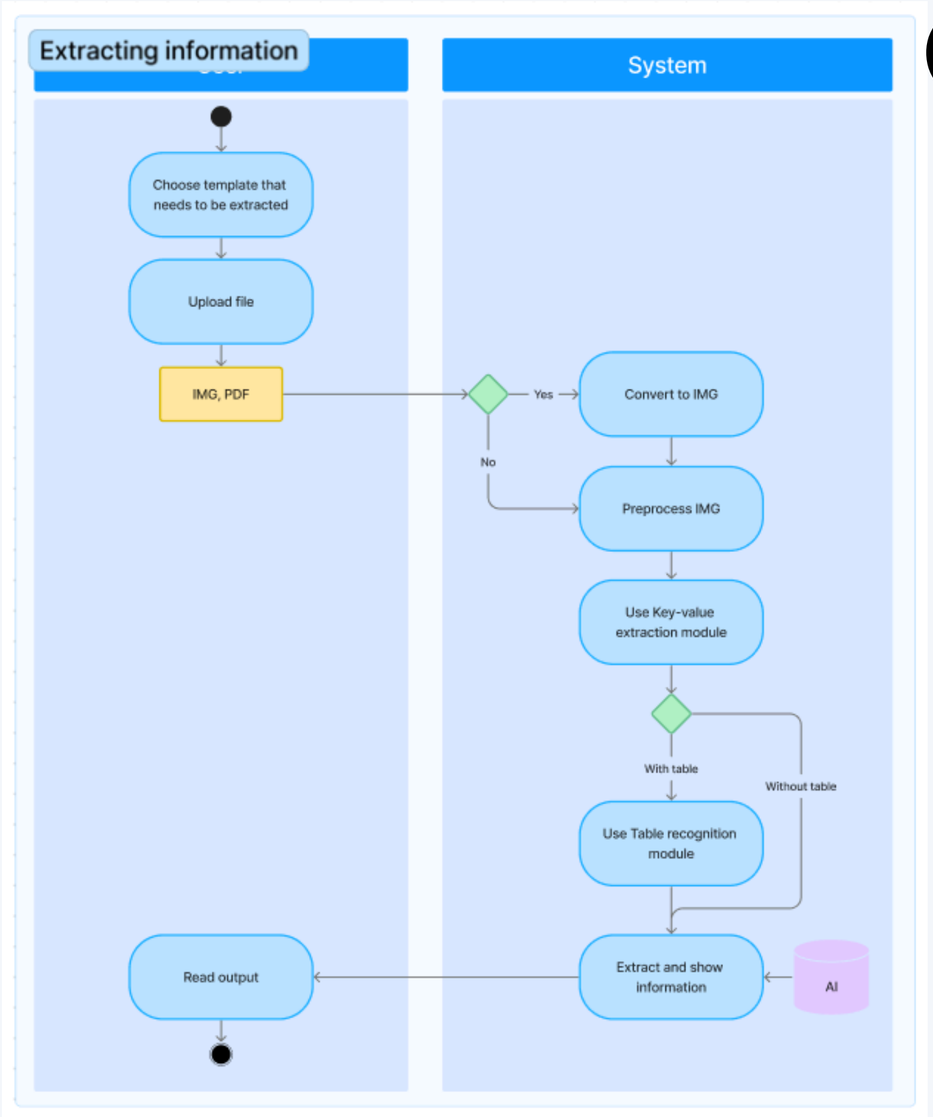
Nguyễn Thanh Hải

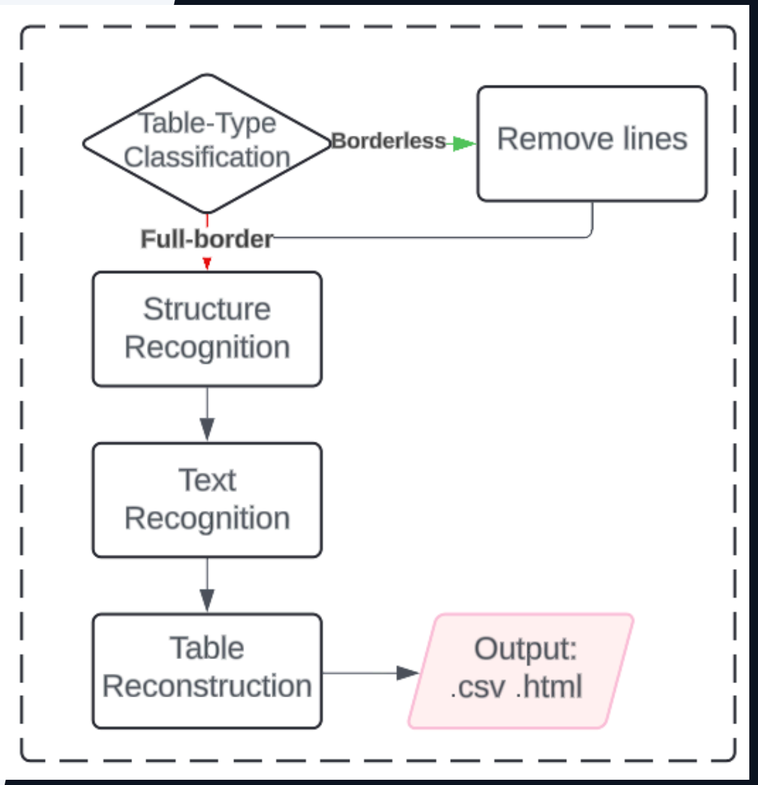
It's hard to gather paper documents that have already been taken by a camera because most of them are personal information, images collected online, and the information will be blurred, so to get enough data, the train model has to prepare its own paper so that the number of images is enough for the model to be effective.

TMA Solutions

Draw Dynamic OCR full-flow

After the demo and feedback from the Product Manager, the workflow for the TR process will have a minor change. If the classification results in a borderless output, it will automatically fill the line and perform reconstruction. The reconstructed version will then be displayed for the user to select and make any necessary corrections, ensuring the template accurately meets the user's requirements.





TMA Solutions

Modify & Optimize TR code

Add Save .htlm .csv

```
import csv
with open('output.csv','w') as result_file:
    wr = csv.writer(result_file, dialect='excel')

    for row, row_text in data.items():
      wr.writerow(row_text)
```

Write an additional 'save' function to save it as an .html file to make it convenient for frontend developers to redraw it on the interface so that the customer can see and confirm that the template requirements have been met.

Add Draw Table from .json

Provide a function to redraw the table, allowing the coders to visually inspect the results before and after fetching the JSON file from the model. This will help verify if the returned data matches the requirements.

```
import pandas as pd
from IPython.display import HTML
import json

with open('data.json', 'r') as f:
    data = json.load(f)
df = pd.DataFrame(data)
html = df.to_html(index=False)
display(HTML(html))
```

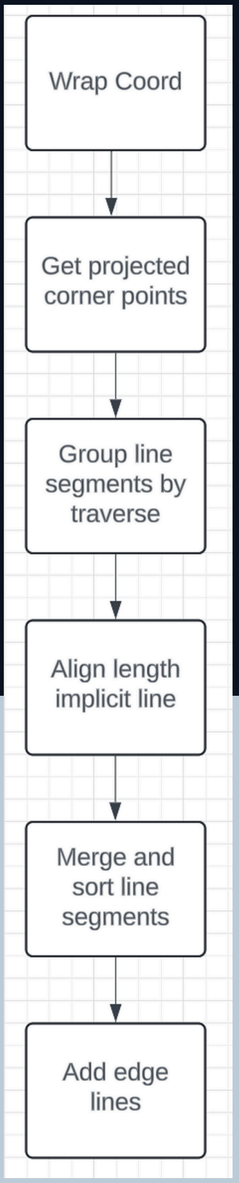
TMA Solutions

Modify & Optimize TR code

Add Crop Cell Image & OCR Text

```
x1 = int(quad[0][0])
y1 = int(quad[0][1])
x2 = int(quad[1][0])
y2 = int(quad[2][1])
cropped_cell = image[y1:y2, x1:x2]
cv2.imwrite(f'cell_{k}.jpg', cropped_cell)
result = reader.readtext(cropped_cell)
text = "\n".join([r[1] for r in result])
text = pytesseract.image_to_string(Image.open(f'cell_{k}.jpg'),
                                   lang='vie')
text_results.append(text)
cells[k]["text"] = text_results
```

The model returns the coordinates of each cell and then uses those coordinates to crop the image of each cell and perform OCR text extraction using pytesseract.



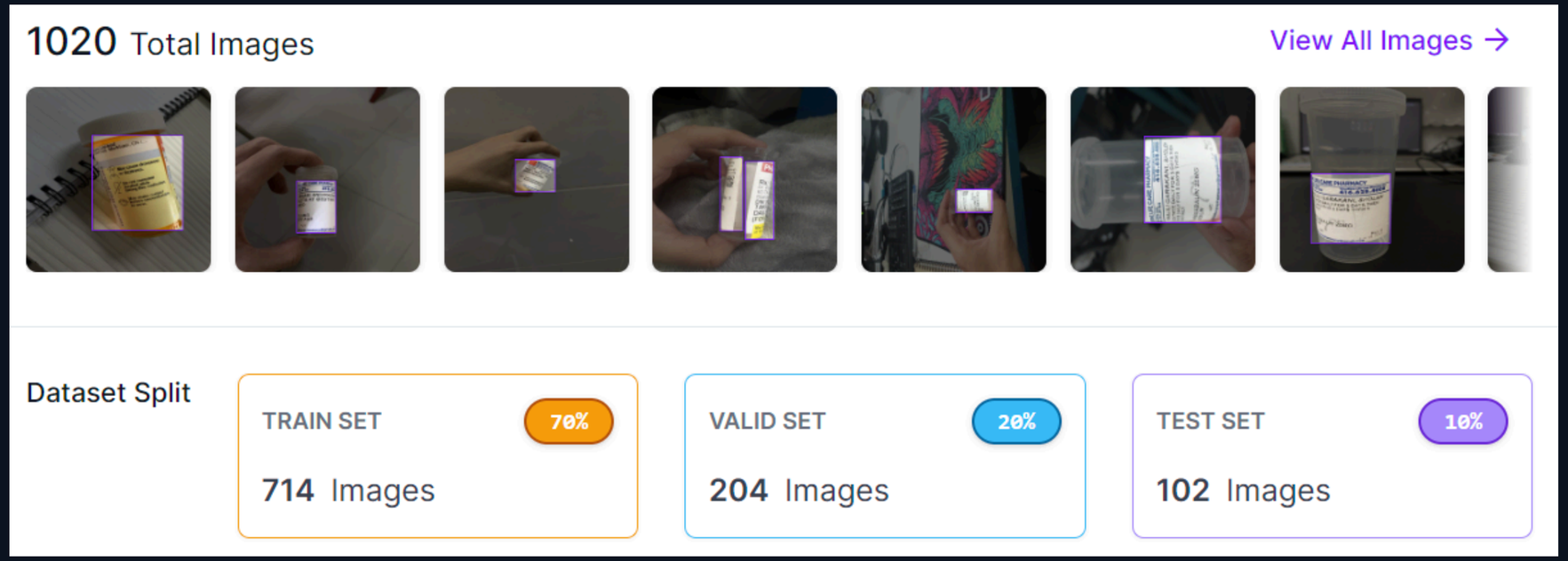
However, adding this functionality makes the model run quite slowly, as it has to process multiple tasks. So I am considering whether I should include this function or just focus on reconstruction and OCR in one pass with key-value pairs.

TMA Solutions

Collect data for training Label Medicine Box

Capture photos of the label information on a total of 17 medication bottles from different angles and backgrounds to obtain 1020 images.

Then, use Roboflow to manually detect the labels in those images, and subsequently train a YOLOv8 model to obtain the best possible model.



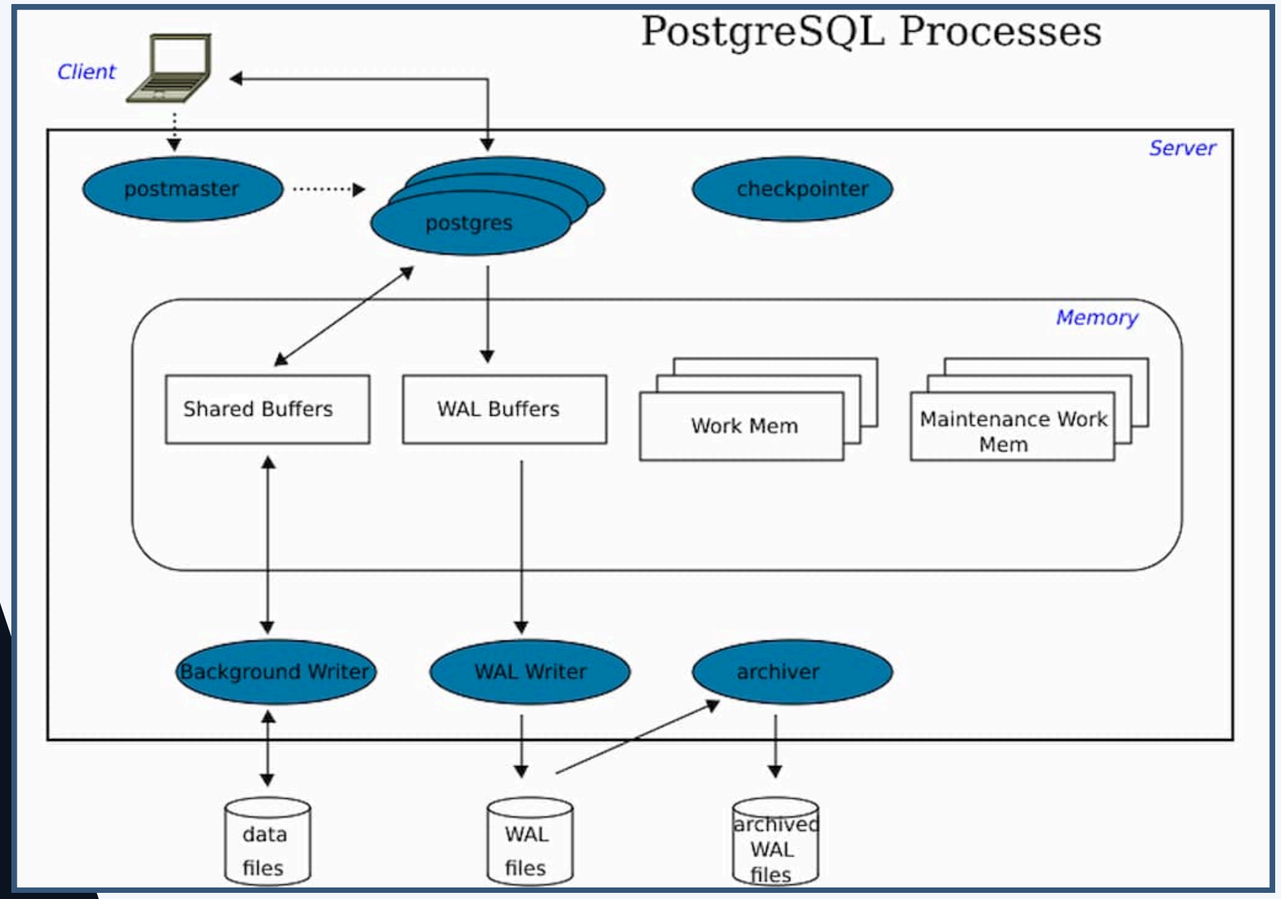

```
Ultralytics YOLOv8.2.2 🚀 Python-3.10.12 torch-2.2.1+cu121 CUDA:0 (Tesla T4, 15102MiB)
Model summary (fused): 168 layers, 11125971 parameters, 0 gradients, 28.4 GFLOPs
Downloading https://ultralytics.com/assets/Arial.ttf to '/root/.config/Ultralytics/Arial.ttf'...
100%|██████████| 755k/755k [00:00<00:00, 16.6MB/s]
val: Scanning /content/drug_bottle_detection-1/valid/labels... 204 images, 0 backgrounds, 0 corrupt: 100%|██████████| 204/204 [00:00<00:00, 1292.70it/s]
val: New cache created: /content/drug_bottle_detection-1/valid/labels.cache

                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100%|██████████| 13/13 [00:09<00:00,  1.38it/s]
                   all        204        212      0.945      0.991      0.992      0.953
Speed: 2.6ms preprocess, 9.3ms inference, 0.0ms loss, 14.2ms postprocess per image
Results saved to runs/detect/val2
```

TMA Solutions

Research on Postgre SQL

PostgreSQL is an open-source, stable database that supports advanced SQL features. It extends SQL with scalable and reliable data handling, making it suitable for diverse applications like mobile, web, geospatial, and analytics.

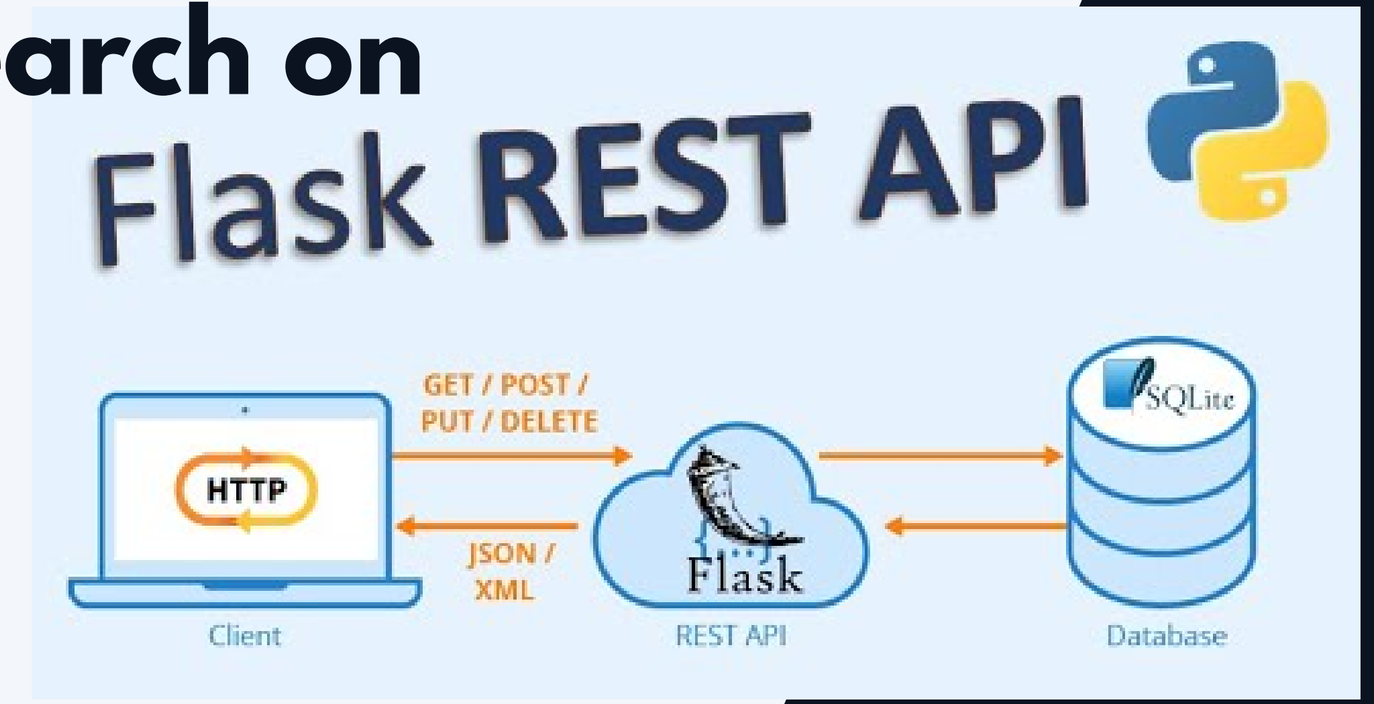


Using object-oriented features in PostgreSQL:

- Communicate with the database servers using objects in their code.
- Define complex custom data types.
- Create functions that work with their own data types.
- Define inheritance, or parent-child relationships, between tables.

TMA Solutions

Research on Flask REST API



Flask API v 2.0

Notes detail

DELETE GET

Retrieve, update or delete note instances.

```
GET http://127.0.0.1:5000/1/
```

```
HTTP 200 OK
Content-Type: application/json

{
    "url": "http://127.0.0.1:5000/1/",
    "text": "build the codez"
}
```

Media type: application/json

Content: {"text": "flask-api is ace"}

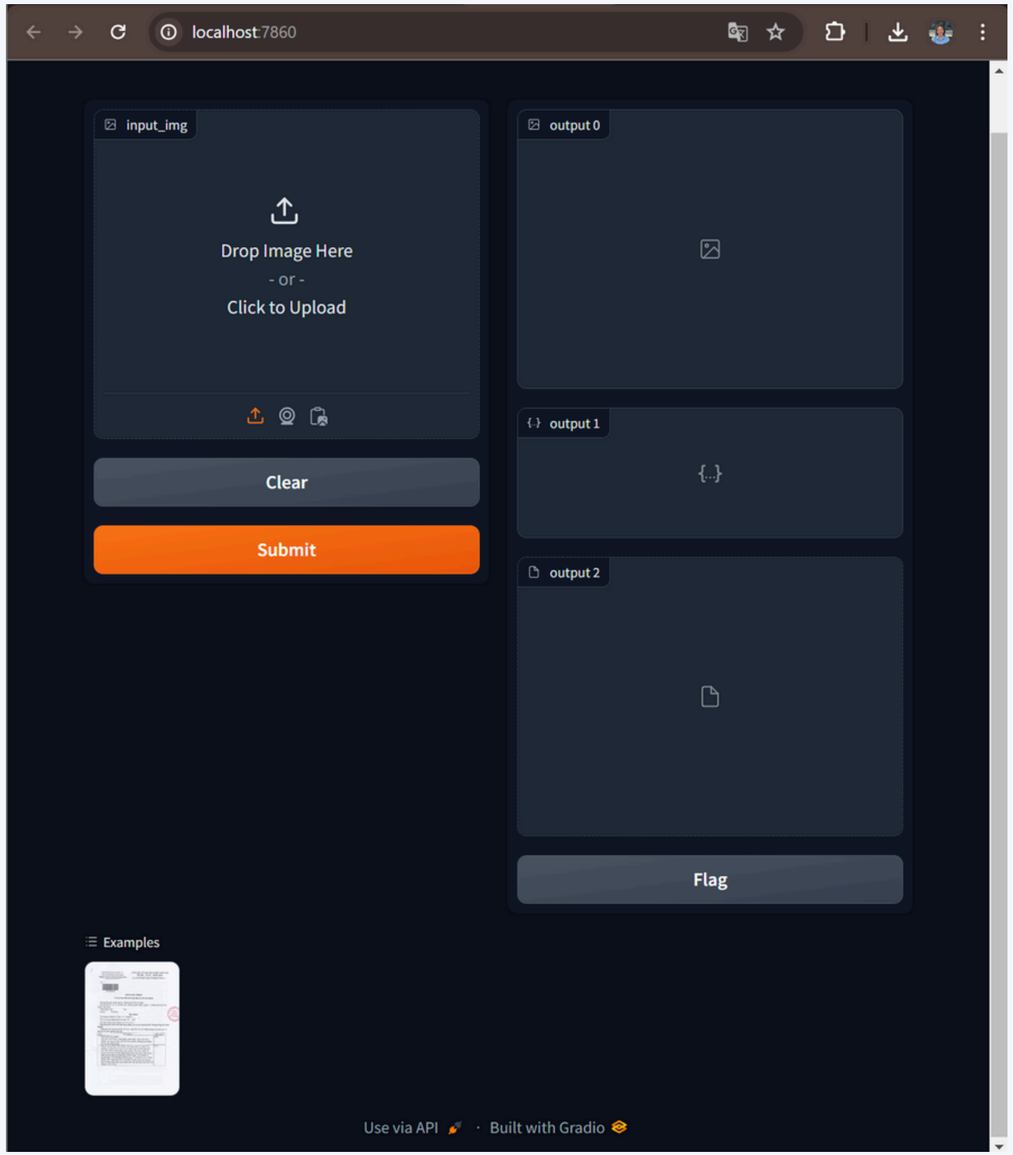
PUT

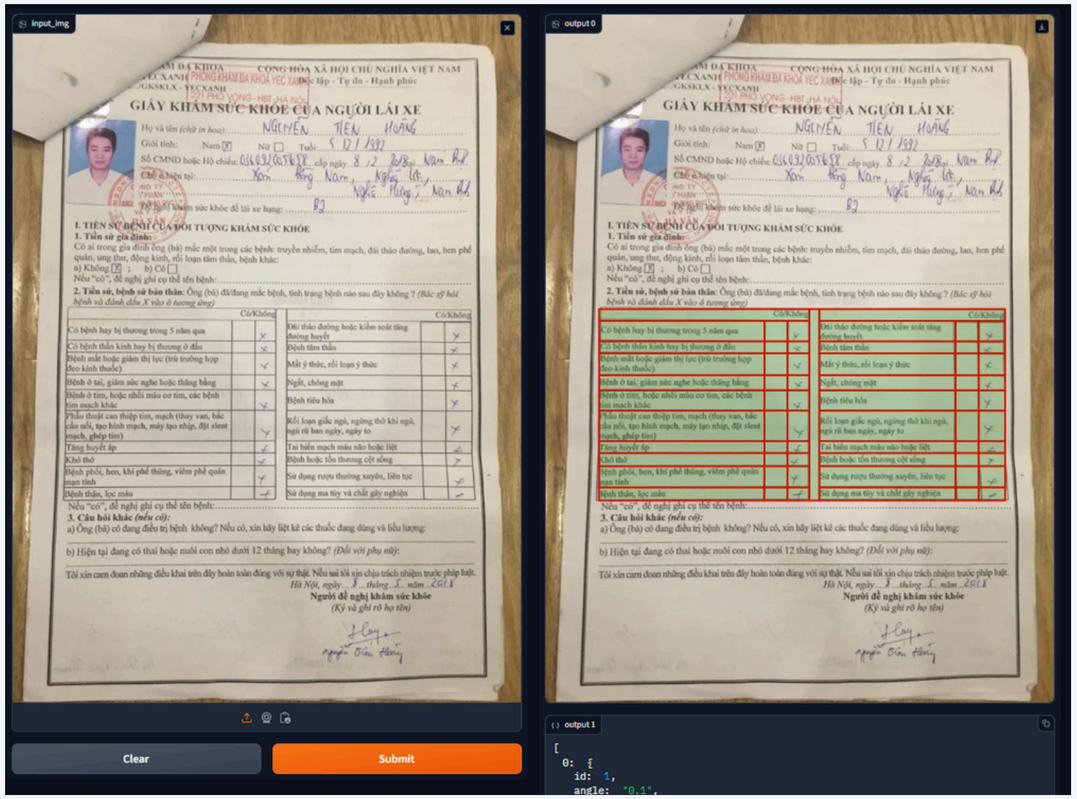
Flask is a well-known framework for web and app development, particularly for creating APIs on the Python platform.

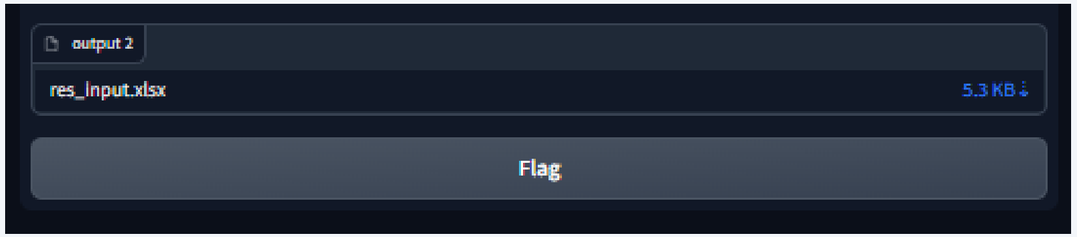
The API is the bridge that transmits data between the client and the server, allowing the application's data to be rendered in an HTML template. The main data interaction methods on the API are GET, POST, PUT, PATCH, and DELETE, but for beginners, understanding GET and POST is typically sufficient.

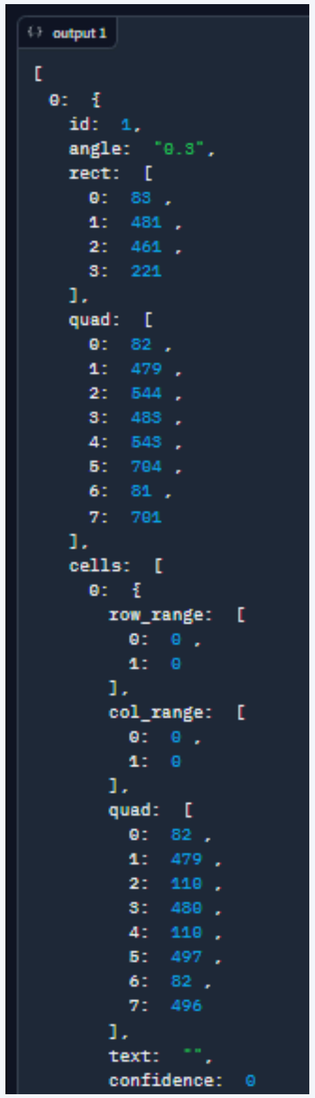
TMA Solutions

Demo TR - TRACE



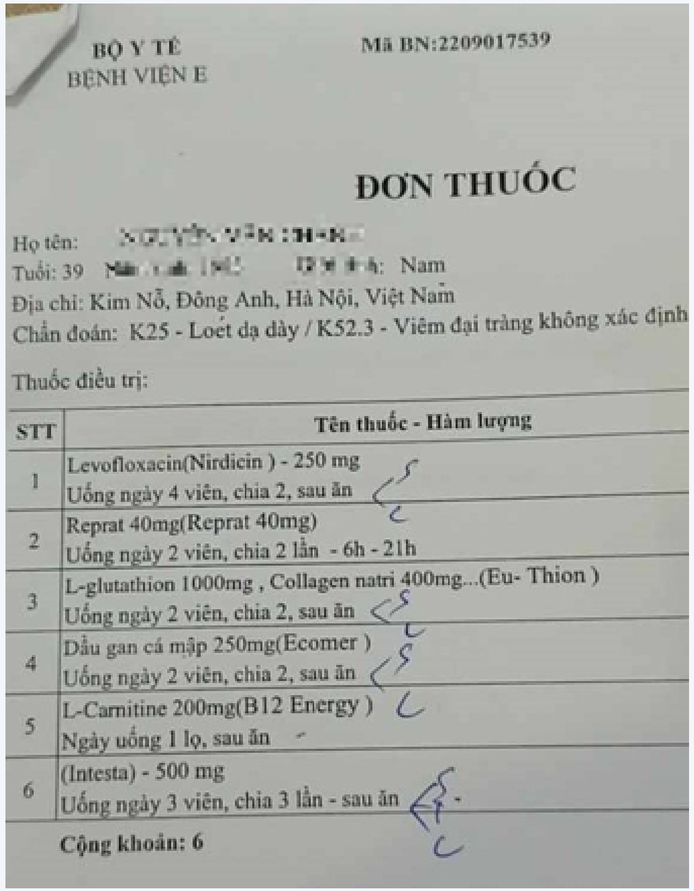






TMA Solutions

Demo TR - Transformer



BỘ Y TẾ
BỆNH VIỆN E

Mã BN:2209017539

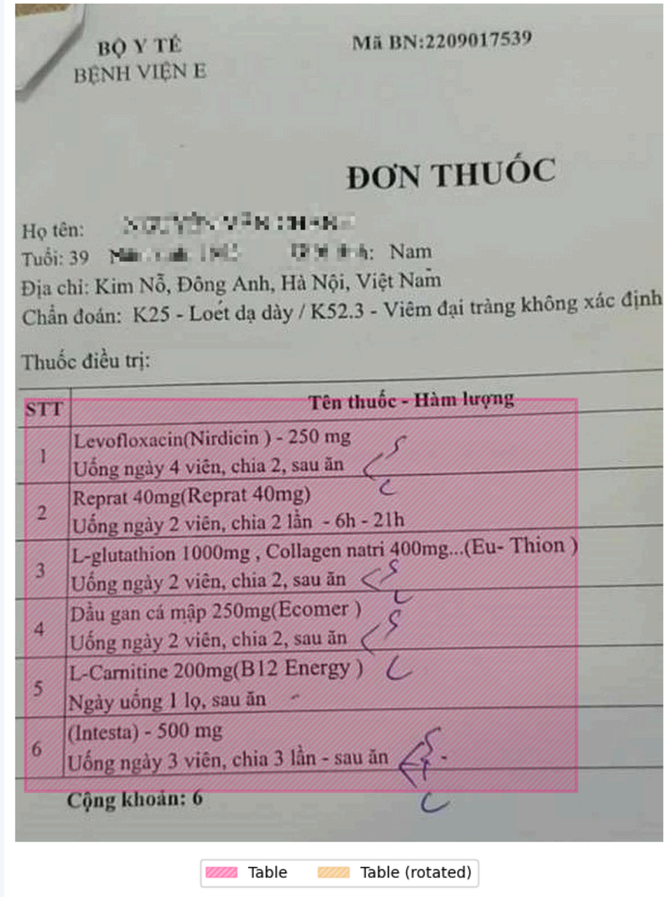
ĐƠN THUỐC

Họ tên: [redacted]
Tuổi: 39 [redacted] Nam
Địa chỉ: Kim Nỗ, Đông Anh, Hà Nội, Việt Nam
Chẩn đoán: K25 - Loét dạ dày / K52.3 - Viêm đại tràng không xác định

Thuốc điều trị:

STT	Tên thuốc - Hàm lượng
1	Levofloxacin(Nirdicin) - 250 mg Uống ngày 4 viên, chia 2, sau ăn
2	Reprat 40mg(Reprat 40mg) Uống ngày 2 viên, chia 2 lần - 6h - 21h
3	L-glutathion 1000mg , Collagen natri 400mg...(Eu- Thion) Uống ngày 2 viên, chia 2, sau ăn
4	Dầu gan cá mập 250mg(Ecomer) Uống ngày 2 viên, chia 2, sau ăn
5	L-Carnitine 200mg(B12 Energy) Ngày uống 1 lọ, sau ăn
6	(Intesta) - 500 mg Uống ngày 3 viên, chia 3 lần - sau ăn

Cộng khoản: 6



BỘ Y TẾ
BỆNH VIỆN E

Mã BN:2209017539

ĐƠN THUỐC

Họ tên: [redacted]
Tuổi: 39 [redacted] Nam
Địa chỉ: Kim Nỗ, Đông Anh, Hà Nội, Việt Nam
Chẩn đoán: K25 - Loét dạ dày / K52.3 - Viêm đại tràng không xác định

Thuốc điều trị:

STT	Tên thuốc - Hàm lượng
1	Levofloxacin(Nirdicin) - 250 mg Uống ngày 4 viên, chia 2, sau ăn
2	Reprat 40mg(Reprat 40mg) Uống ngày 2 viên, chia 2 lần - 6h - 21h
3	L-glutathion 1000mg , Collagen natri 400mg...(Eu- Thion) Uống ngày 2 viên, chia 2, sau ăn
4	Dầu gan cá mập 250mg(Ecomer) Uống ngày 2 viên, chia 2, sau ăn
5	L-Carnitine 200mg(B12 Energy) Ngày uống 1 lọ, sau ăn
6	(Intesta) - 500 mg Uống ngày 3 viên, chia 3 lần - sau ăn

Cộng khoản: 6

Table Table (rotated)

STT	Tên thuốc - Hàm lượng
1	Levofloxacin(Nirdicin) - 250 mg Uống ngày 4 viên, chia 2, sau ăn
2	Reprat 40mg(Reprat 40mg) Uống ngày 2 viên, chia 2 lần - 6h - 21h
3	L-glutathion 1000mg , Collagen natri 400mg...(Eu- Thion) Uống ngày 2 viên, chia 2, sau ăn
4	Dầu gan cá mập 250mg(Ecomer) Uống ngày 2 viên, chia 2, sau ăn
5	L-Carnitine 200mg(B12 Energy) Ngày uống 1 lọ, sau ăn
6	(Intesta) - 500 mg Uống ngày 3 viên, chia 3 lần - sau ăn

	STT	Ien (uuC
0	Levofloxacin(Nirdicin) - 250 mg [Uóng npày 4 ...	NaN
1	Reprat 4Omg(Reprat 40mg) [Uóng ngày 2 vienchia...	NaN
2	[L-glutathion IOOOmg , Collagen natri 4OOmg- (...	NaN
3	[Dẩu gan cá mp 250mg(Ecomer luóng ngày 2 viên;...	NaN
4	L-Camitinc 200mg(B12 Energy [Ngày uong 1o, sau ún	NaN
5	(Intesta) 500 Uống ngẩy 3 viên; chia 3 lan sa...	NaN

```
Max number of columns: 2
['STT', 'Ien (uuC']
['Levofloxacin(Nirdicin ) - 250 mg [Uóng npày 4 vién, chia 2. sau ẳn', '']
['Reprat 4Omg(Reprat 40mg) [Uóng ngày 2 vienchia 2 làn 6h - 2Ih', '']
['[L-glutathion IOOOmg , Collagen natri 4OOmg- (Eu- Thion Uông ngày 2 viên, chia 2 sau &n', '']
['[Dẩu gan cá mp 250mg(Ecomer luóng ngày 2 viên; chia 2, sau &n', '']
['L-Camitinc 200mg(B12 Energy [Ngày uong 1o, sau ún', '']
['(Intesta ) 500 Uống ngẩy 3 viên; chia 3 lan sau án mg', '']
```

TMA Solutions

Review Demo TRACE

Advantages:

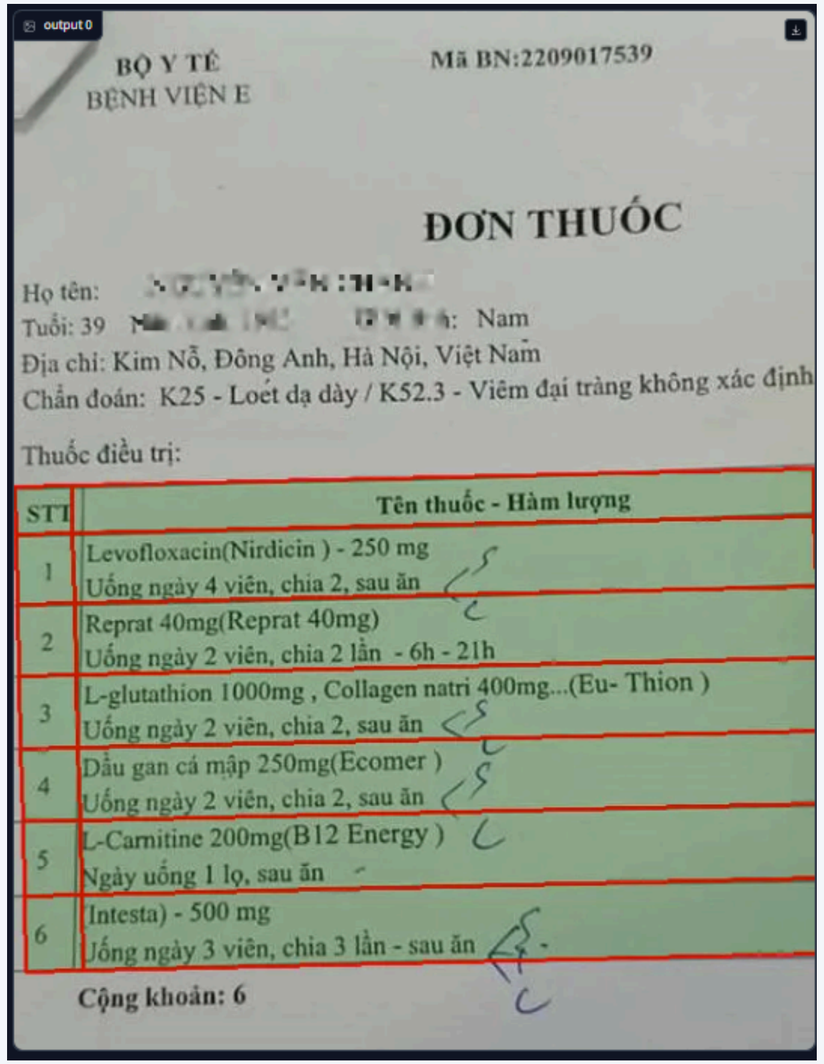
- Fast, Quite high accuracy, Can detect span-cells
- Separate modules for two classes - fullborder/ borderless

Disadvantages:

- Even for borderless tables, it reconstructs them as fullborder
- Tends to merge two close tables into one
- Requires CUDA usage

Comparison to other modules:

- Faster and more accurate than modules like PaddleOCR-Transformer
- GPU usage is mandatory



TMA Solutions

Lessons learnt

I have learned the following skills:

- How to create a plan for the week and for each sprint.
- How to fill out a daily report to ensure the progress of work.
- How to share updates during each sprint review.
- How to work effectively in a team and ensure the efficiency and results of my own tasks.
- The process of researching a new topic and implementing it according to the project requirements.
- How to use GitLab to integrate the team's code.
- How to support each other to ensure the project's progress.
- Gained a lot of new knowledge from the mentor PM and colleagues.
- Received training on various topics such as the product development process and teamwork.

These skills have helped me develop a structured approach to project management, collaboration, and continuous learning to meet the demands of the project.

Best practices

I have improved the following skills:

- Planning and executing tasks more effectively within the project.
- Enhancing my technical expertise and domain knowledge.
- Improving my ability to share problems and project updates with others.
- Becoming more extroverted, friendly, and caring towards my colleagues.
- Enhancing my English proficiency by working in an English-speaking environment.
- Consistently meeting deadlines and ensuring timely completion of assigned tasks.
- Being more proactive in communicating my issues to supervisors and co-workers to receive timely assistance.

The improvements I've made have made me a more effective and valuable team member, greatly contributing to the project's success.

TMA Solutions

Thanks

Thank you to the company for providing me the opportunity to intern in the field I'm passionate about pursuing. I'm grateful to the interns in the department who have always supported us. Their help enabled me to complete the internship in the best possible way - this has been a very meaningful period for me.

I would like to express my sincere gratitude to Nguyen Minh Nghia and Tran Trung Kien for guiding and supporting me, for giving me the chance to gain experience, develop myself, and enhance my domain knowledge.
I'm also thankful to my colleagues who have been by my side, supporting each other in our work and personal lives.