โครงสร้าง Project

/OCR-Lanna

```
├── dataset/              # โฟลเดอร์เก็บชุดข้อมูล
│   ├── model/
│   ├── text/
│   ├── text binary/
│   ├── ตัวอักษรล้านนา
│   ├── ใบลาน
├── image_for_test/       # โฟลเดอร์เก็บรูปภาพสำหรับทดสอบ
├── Model/                # โฟลเดอร์เก็บโมเดลที่เทรนแล้ว
├── template/             # โฟลเดอร์เก็บไฟล์เทมเพลต (HTML)
├── API.py                # โค้ดสำหรับ API ของระบบ
├── Predict.py            # โค้ดสำหรับทำนายผล
├── track.py              # โค้ดสำหรับตรวจหาอักษรใน dataset
├── Train_model.py        # โค้ดสำหรับฝึกโมเดล
├── requirements.txt      # ไฟล์ระบุแพ็กเกจที่ต้องติดตั้ง
```

วิธีการใช้งาน API

- ให้รันไฟล์ API.py จากนั้นหน้าเว็บไซต์จะขึ้นมาให้อัพโหลดรูปภาพ

วิธีการ Train model ใหม่

- หากเริ่ม Train ใหม่ ให้เริ่มจากเก็บ dataset เพิ่มก่อน ต้องรันไฟล์ track.py เพื่อตรวจจับตัวอักษรทุกตัวใน dataset
- หากรันเสร็จแล้ว รูปตัวอักษรทุกรูปจะอยู่ใน dataset/text
- คัดแยกตัวอักษรเพิ่มจาก dataset/text นำมาแยกไว้ที่ dataset/model/train/..

  ใน ../train จะมี Folder ชื่อภาษาอังกฤษตามตัวอักษร ใช้ภาษาอังกฤษเนื่องจากโมเดลไม่สามารถอ่าน target ที่เป็นอักษรไทยได้

  ดังนั้นแต่ละตัวอักษรจะถูกแปลงเพื่อใช้สำหรับ Train model ดังนี้

Dictionary :

_n : –น	N : น	H : ห
_m : –ม	B : บ	HL : หลฯ
_o : –อ	PA : ป	OY : อยฯ
K : ก	PH : ผ	A : ะ
KH : ข	F : ฝ	Aa : ั
C : ค	P : พ	AAA : า
NG : ง	M : ม	EI : ิ
J : จ	Y : ย	EE : ี
CH : ฉ	R : ร	EE : ี
NN : ณ	L : ล	EU : ุ
D : ด	V : ว	EA : เ
T : ต	S : ส	AI : ไ

“ – ” คือพยัญชนะที่เป็นตัวสะกด | “ ฯ ” คืออักษรควบกล้ำ

- สามารถเพิ่มพยัญชนะโดยสร้าง folder กำหนดตัวอักษรภาษาอังกฤษที่ไม่ซ้ำกันและไปเพิ่มพยัญชนะใน Dictionary ส่วนบนของ Code ทุกไฟล์
- รันไฟล์ Train_model.py เพื่องทำการ Train Model
- หลังจากรันเสร็จ หากไม่มี error แจ้ง model จะถูกบันทึกใน Folder Model/
- สามารถรันไฟล์ API.py หรือ อื่นๆเพื่อทดสอบโมเดล