

Thai Nutrition Table Extraction

[Thai-Nutrition-Table-Extraction on GitHub](#)

This project is created to extract data form Nutrition Table to help user collect data about nutrition to use with Health care application, Fitness app, etc.

NOTIC Thank to [text-detection-ctpn](#) and [Tesseract](#), we using their source code for detect and recognize text in nutrition table.

Requirement

- Computer running **Linux** or **MacOS**
- **Python 3.7.1** or later
- **Pip 20.1** or later

Setup

1. Install Python libraries.

```
pip install -r requirements.txt
```

2. Check directory `text_detection/checkpoints_mlt`. If directory not exists, download the file from [google drive](#) or [baidu yun](#). Then extract file and put `checkpoints_mlt/` in `text-detection/`.
3. Setup `nms` and `bbox`. Because of the libraries are written in Cython, hence you have to build the library by using follow command.

```
cd text_detection/utils/bbox
chmod +x make.sh
./make.sh
```

4. [Install Tesseract](#) by following [this document](#).
5. Install Tesseract pretrained to supporting Thai language by going to [this page](#) and download `tha.traineddata`. Then set the `TESSDATA_PREFIX` environment variable and put file in `TESSDATA_PREFIX/tessdata/tha.traineddata`.

Demo

- Run `main.py` to see result.

```
python main.py
```