Image to Text with Tesseract-OCR, Python, Flask

Step1: Create virtual environments

Go to home

$ cd

$ mkdir tuts

$ cd tuts

Create virtual environments : Flask/

$ python3 -m venv Flask/

$ source ./Flask/bin/activate

$ cd Flask/

Step 2: Install Flask

$ pip3 install Flask
$ python3 -m flask –version

Step 3: Environment Variables

Create new folder for the app.
$ mkdir img2txt
$ cd img2txt

Open a new file: .flaskenv
$ vim .flaskenv

And add the content into .flaskenv:

```
FLASK_APP=app.py              # or whatever your app name
FLASK_ENV=development     # or production
```

Step 3: Testing

# In Flask, create a file app.py
$ vim app.py

Then add the following to app.py

```
"""
from flask import Flask
app = Flask(__name__)
```

```
@app.route('/')
def hello_world():
    return "Hello World!"
"""

# Testing
$ flask run
```

And check your browser at http://127.0.0.1:5000/

OK! Next we'll build an app for image-to-text

Step 4: Install Dependencies

Plz refer to my Github for source code
(Github link in the description below)

```
pip3 install python-dotenv
pip3 install numpy
pip3 install opencv-python
pip3 install pytesseract

sudo apt update
sudo apt install tesseract-ocr
sudo apt install libtesseract-dev
```

Step 5: Run the app locally

Plz refer to my Github for source code
(Github link in the description below)

Directory structure is shown here =======>>>>

```
img2txt
    static/
    templates/
    app.py
    Aptfile
    requirements.txt
    ...

$ flask run
```

And check the browser on:
http://127.0.0.1:5000/

Step 6: Deploy on Heroku

Read guide.txt for details!

Important files:
1. app.py : the application
   (refer to my Github for the source code)
2. Aptfile : the third-party dependencies for Heroku

to install (e.g: tesseract-ocr)
3. Procfile : a list of process types in an app
(on Heroku)
4. requirements.txt : a list of dependencies to install
5. runtime.txt : version of Python to run on Heroku

Note:
Add this line to app.py
pytesseract.pytesseract.tesseract_cmd =
"/app/.apt/usr/bin/tesseract"
(refer to my Github for the source code)

> Login to Heroku, and create a new app:
heroku login

> go to the app directory
cd <path/to/>/img2txt (e.g: cd tuts/Flask/img2txt)
git init

heroku create img2txt-ocr
heroku git:remote -a img2txt-ocr

> Add 02 Buildpacks:
1. https://github.com/heroku/heroku-buildpack-apt
2. heroku/python
heroku buildpacks:add --index 1 https://github.com/heroku/heroku-buildpack-apt
heroku buildpacks:add --index 2 heroku/python

> Add Config Vars:
TESSDATA_PREFIX = ./.apt/usr/share/tesseract-ocr/4.00/tessdata
heroku config:set TESSDATA_PREFIX=/app/.apt/usr/share/tesseract-ocr/4.00/tessdata

> heroku stack (heroku-20) has bad compatibility with tesseract.
You may need to change heroku stack from 20 to 18 using command:
heroku stack:set heroku-18

> Deploy app on Heroku
git add .
git commit -m "Deploy app on Heroku"

heroku git:remote -a img2txt-ocr
git push heroku master

(Check here on my Heroku: https://img2txt-ocr.herokuapp.com/)

> Check logs
heroku logs --tail

> Login to Heroku CLI
heroku run bash
which tesseract

find -iname tessdata

DONE!
THANK YOU!