



EXPLORER

V TENCENT_OCR

_pycache

>

lprof

venv

d

0

.env

.env.example

.gitignore

demo_ocr_api.py

X

ocr_call.py

output.pdf

README.md

2reconstruct_pdf.py

= requirements.txt

□

demo_ocr_api.py

ocr_call.py

reconstruct_pdf.py X

2

pagesizes.py

.gitignore

.env.example

0

□

v

M

reconstruct_pdf.py;

1

import json

2 from reportlab.pdfgen import canvas

3 from reportlab.lib.pagesizes import A3

4 from ocr_call import get_ocr_results # Import the function from your ocr_api.py

5

Call the function to get oCR results

7 ocr_data =get_ocr_results("http://public.yuantsy.com/PDF_Split/CS188/31.pdf")

8 ocr_result = json.loads(ocr_data)

C

10

Create a PDF canvas

11 c = canvas.Canvas("output.pdf",pageize=A3)

12

width, height = A3 # Get page size

13

print(width, height)

14

15 # Extract TextDetections if it exists in the response

16

if 'TextDetections' in ocr_result:

17

for item in ocr_result['TextDetections']:

18

text = item['DetectedText']

19

X= item['ItemPolygon']['X']

20

=heiaht - item['ItemPolygon']['Y'] # Adjust Y coordinate

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

Python

■v

X

{}){"Language": "zh","Angle":359.989990234375,"PdfPageSize":1,"RequestId":"97a5927b-9273-41a6-b54a-32fc444c4f18"}

(.venv) /Users/chenyiyun/Desktop/tencent_ocr/.venv/bin/python /Users/chenyiyun/Desktop/tencent_ocr/reconstruct_pdf2.py

/Users/chenyiyun/Desktop/tencent_ocr/.venv/lib/python3.11/site-packages/requests/__init__.py:102: RequestsDependencyWarning: urllib3 (2

2.1) or chardet (5.2.0)/charset_normalizer (3.3.2) doesn't match a supported version!

warnings.warn("urllib3 ({}), or chardet ({}), or charset_normalizer ({}), doesn't match a supported

(.venv) /Users/chenyiyun/Desktop/tencent_ocr/.venv/bin/python /Users/chenyiyun/Desktop/tencent_ocr/reconstruct_pdf.py

/Users/chenyiyun/Desktop/tencent_ocr/.venv/lib/python3.11/site-packages/requests/__init__.py:102:RequestsDependencyWarning: urllib3 (2

2.1) or chardet (5.2.0)/charset_normalizer (3.3.2) doesn't match a supported version!

warnings.warn("urllib3 ({}), or chardet ({}), or charset_normalizer ({}), doesn't match a supported

841.8897637795277 1190.5511811023623

Traceback (most recent call last):

File "/Users/chenyiyun/Desktop/tencent_ocr/reconstruct_pdf.py", line 27, in <module>

c.moveTo(points[0][0],points[0][1])

AttributeError: 'Canvas' object has no attribute 'moveTo'

.venv

OUTLINE

TIMELINE

><

80A0

Ln8,Col34

Spaces:4

UTF-8

LF

()Python

3.11.6 (.venv: venv)

28

8

0