Table of Contents

OCR a PDF file 2

OCR a PDF file

Super simple solution.

1. Install `tesseract` and the corresponding data package for your language
2. Install `ocrmypdf` either from AUR or pip.

If your document is scanned/has no text layer, run it as

```
ocrmypdf input.pdf output_with_text.pdf
```

If it does have text, it doesn't always merge cleanly. You can use `--force-ocr` but that rasterizes the file and makes it massive.

From:
https://wiki.tonytascioglu.com/ - **Tony Tascioglu Wiki**

Permanent link:
https://wiki.tonytascioglu.com/scripts/files/ocr_pdf_file

Last update: **2023-03-27 02:21**