

[1]: <https://github.com/ocrmypdf/OCRmyPDF> "" [2]: <https://pdf.iskysoft.com/ocr-pdf/open-source-ocr.html> "" [3]: <https://source.opennews.org/articles/our-search-best-ocr-tool-2023/> "" [4]: <https://ocr.space/searchablepdf> ""

Certainly! To create a searchable PDF from scanned images while preserving the original look of the pages, you can use an open-source tool called **OCRmyPDF**. It's designed to add an OCR text layer beneath the scanned images, allowing the text to be searchable without altering the visual presentation of the pages.

Here's a step-by-step guide on how to do it:

1. Install OCRmyPDF:

- For Linux: Use the package manager, for example, `apt install ocrmypdf` for Debian/Ubuntu.
- For Windows: You can use the Windows Subsystem for Linux (WSL) and then install it with `apt install ocrmypdf`.
- For macOS: Install using Homebrew with `brew install ocrmypdf`.

2. Run OCRmyPDF:

- Open a terminal or command prompt.
- Use the following command to process your PDF file:

```
bash ocrmypdf --input-language eng --rotate-pages --deskew --output-type pdfa input_scanned.pdf output_searchable.pdf
```
- Replace `input_scanned.pdf` with the path to your scanned PDF file.
- The `output_searchable.pdf` will be the new searchable PDF file.

3. Options Explained:

- `--input-language eng` : Specifies the language for OCR; use `eng+fra` for English and French, for example.
- `--rotate-pages` : Automatically corrects pages that are misrotated.
- `--deskew` : Straightens pages that are crooked.
- `--output-type pdfa` : Ensures the output is PDF/A, a format for long-term archiving.

OCRmyPDF uses the Tesseract OCR engine, which can recognize more than 100 languages, and it's capable of handling files with thousands of pages!^[1].

Remember to check the documentation of OCRmyPDF for more detailed instructions and additional options that you might find useful for your specific needs.

Source: Conversation with Bing, 4/26/2024 (1) OCRmyPDF adds an OCR text layer to scanned PDF files, allowing them to

<https://github.com/ocrmypdf/OCRmyPDF>. (2) Top 3 Open Source OCR Software - iSkysoft PDF Editor. <https://pdf.iskysoft.com/ocr-pdf/open-source-ocr.html>. (3) Our search for the best OCR tool in 2023, and what we found. <https://source.opennews.org/articles/our-search-best-ocr-tool-2023/>. (4) PDF OCR: Create Searchable PDF (Sandwich PDF) from Scan online for free. <https://ocr.space/searchablepdf>.