

# R Shiny OCR PDF tool

This OCR (Optical character recognition) Shiny app is to convert PDF content into text to assist data entry or text analysis. The App is for generic usage. It works well with plain text from good quality scans. It works with English language by default, but more languages can apply with Tesseract language settings.

User guide,

1. Upload PDF from the input area on the top right.
2. Adjust the page width for better page resolution. The right resolution will also effect on the text recognition quality. Choose a bigger width if the text is small.
3. Rotate the page if required, the rotation is to the right side, unit by degree.
4. User can choose the page you want.
5. User can click and drag on the PDF page, to crop the area of interest, the extract text will be generated on the left. Note, margins are recommended on the image cropping. Different font and font size may affect the extraction quality.
6. Alternatively, user can click on the "Extract full page" to extract the text on the page, however, the outcome usually needs further processing as the OCR doesn't read the layout.
7. Uploading a new PDF file will reset the page selections.