

# tesseract i ocrmypdf - Instalacija - Windows

written by js | 2022-08-28

## Preuzmi i instaliraj

- tesseract: <https://github.com/UB-Mannheim/tesseract/wiki> (instaliraj sve module za željene jezike)
- ocrmypdf: <https://ocrmypdf.readthedocs.io/en/latest/installation.html#native-windows>

## Dodatne podesidbe

- Preuzmi novi language pack za sve jezike odabrane tijekom instalacije (npr. "hrv.traineddata") - <https://github.com/tesseract-ocr/tessdata/>
- Spremi u "C:\Program Files\Tesseract-OCR\tessdata\"
- Opcija: `preuzmi jbig2.exe` - <https://github.com/anotatta/jbig2enc/releases/tag/0.29>
- Opcija: spremi jbig2.exe u folder "C:\Program Files\Tesseract-OCR\"
- Dodaj "C:\Program Files\Tesseract-OCR\" u PATH

## Oceeriraj

Otvori novi CMD i ocriraj nešto:

```
ocrmypdf --language hrv+eng neštonešto.pdf neštonešto-ocr.pdf  
--force-ocr --sidecar neštonešto-ocr.pdf.txt
```

Ako je sve okej, po svršetku ocriranja ćeš imati datoteku "neštonešto-ocr.pdf" koja je pretraživa (srčabilna) i datoteku "neštonešto-ocr.pdf.txt" koja je - tekst, jelte.

---

ocr xocr xtesseract xocrmypdf xpdf2txt pdf2txt pdf2pdf