

# Manipulacija PDF datotekama {CLI}


written by js | 2021-02-25

## Tema

Rukovanje PDF datotekama. Pretvaranje u tekst, rezanje, spajanje, odvajanje stranica, ...











---

## NAPS2 - Not Another PDF scanner 2

Recommended download:  [naps2-6.1.2-setup.exe](#)

Or, choose one of the files below:

### Latest release

OS	Version	Type	DL	Size
	6.1.2	Installer	 <a href="#">exe</a>	2.27 MB
	6.1.2	Installer for group policy	 <a href="#">msi</a>	2.53 MB
 	6.1.2	Portable archive	 <a href="#">zip</a>	2.84 MB
 	6.1.2	Portable archive	 <a href="#">7z</a>	1.91 MB

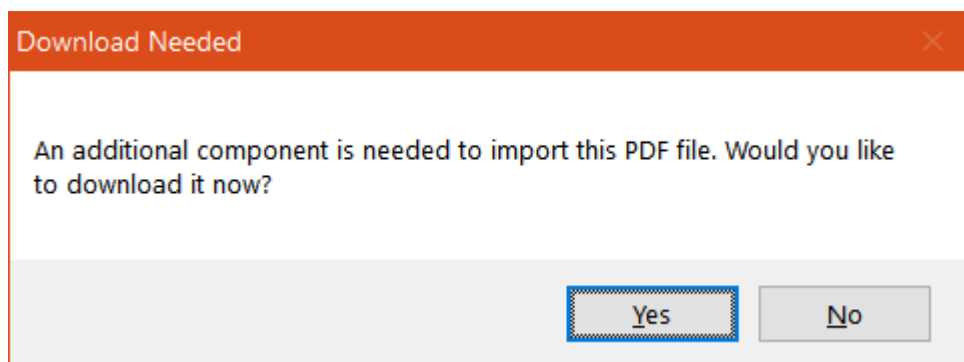
[Last updated July 27, 2019](#) · [Changelog](#) · [Hashes](#)

<https://www.naps2.com/download.html>

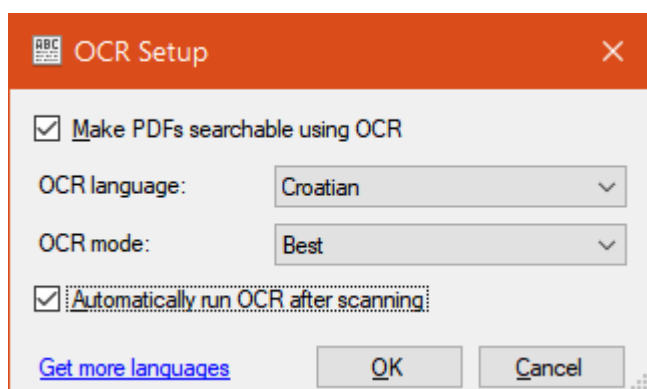
Ako ste na svojim računalima - preuzmite "Recommended download" i instalirajte.

Ako ste na ne-svojim računalima (tipa u uredu gdje nemate administrativne ovlasti) - preuzmite jednu od "Portable archive", zip ili 7z.

Nakon instalacije ili *anzipanja* pokrenite program, otvorite neki PDF i dopustite da preuzme dodatne module.



Pa kliknite na "OCR", odaberite Croatian i pustite da preuzme "rječnik".



Sad možete skenirati (ili *importirati*) PDF.

*Importirajte* više PDF-ova i premećite stranice kako želite (drek & drop).

Označite neke stranice i spremite ih.

(*source*, za oni koji misle da je itko ikada pročitao ičiji tuđi *source-code*:  
<https://github.com/cyanfish/naps2>)

---

## Razni alati (Linux, WSL)

- `sudo apt install pdf-redact-tools pdf2djvu pdfarranger pdfchain pdfmod pdfposter pdftk pdfgrep`

- `pdftopdf -i "trazim ovaj tekst" dokument.pdf # traži tekst`
- `pdftopdf --ignore-case ".*" dokument.pdf # praktički "pdf2txt"`

---

## Apache PDF box (Java)

Download: <https://pdfbox.apache.org/download.cgi>

### Izvuci tekst iz PDF datoteke (PDF2TXT):

- `java -jar pdfbox-app-X.Y.ZZ.jar ExtractText dokument.pdf dokument.txt`

### PDF2TXT - novije verzije PDFBoxa:

- `java -jar pdfbox-app-X.Y.ZZ.jar export:text --input=1.pdf --output=1.txt -encoding=UTF-8`

(pazi na velika/mala slova, broj crtica "-" kod argumenata, kaos..)

### Podijeli (" split ") dokument

Podijeli X.PDF ("-i X.pdf") na više dokumenata naziva "xSPLIT-1.pdf" i "xSPLIT-2.pdf" ("-outputPrefix=xSPLIT"), dijelivši nakon svake 2. stranice ("-split=2"):

- `java -jar pdfbox.jar split -i X.pdf -outputPrefix=xSPLIT -split=2`

Izvuci stranice 3-6 iz X.PDF:

- `java -jar pdfbox.jar split -i X.pdf -outputPrefix=xSPLIT -startPage=3 -endPage=6`

### **Spoji dva dokumenta ("1.pdf" i "2.pdf") u jedan ("3.pdf"):**

- `java -jar pdfbox-app-X.Y.ZZ.jar merge --input=1.pdf --input=2.pdf --output=3.pdf`

### **Podijeli PDF (izvuci stranice od 21 do 31):**

- `java -jar e:\ut\pdfbox.jar split -i 01.pdf -outputPrefix=str-21-31 -startPage=21 -endPage=31`

<https://pdfbox.apache.org/1.8/commandline.html>

---

xpdfx xhackx xhakiranjex xpdf2txt xpdftotxt xpdf2text xpdftotext pdf2txt  
pdftotxt pdf to text pdf2text pdftotext xclix xbashx xcmtx hx hxhx  
NAPS2.Portable.exe xNAPS2.Portable.exe