Brugsanvisning: OCR-konvertering af billedfiler til PDF

Dette værktøj laver automatisk OCR-behandlede PDF-filer ud fra billedfiler (TIFF, JPG, JPEG, GIF). Brug det, når du har scannede avisudklip eller lignende små artikler, du vil kunne søge i.

Er der flere billeder der skal samles til én pdf-fil, er dette script ikke egnet.

Forudsætninger

- **Programmet Tesseract** skal være installeret på computeren og tilføjet til systemets `PATH`
- Download her: https://sourceforge.net/projects/tesseract-ocr.mirror/
  Husk at tilføje Danish og German samt Fraktur.
- Scriptet ligger i:
  `M:\Arkivet generelt\Edb og progammer tips\Programmer\Scripts\ocr_billeder.bat`

Sådan bruger du scriptet

1. **Dobbeltklik** på `ocr_billeder.bat` i mappen/skrivebord
2. Du bliver bedt om at **indtaste stien** til den mappe, der indeholder dine billedfiler (f.eks. `O:\U - Udklip`).
3. Du bliver spurgt, om **pc'en skal slukkes**, når behandlingen er færdig.
   - Skriv `ja` hvis du ønsker det — ellers tryk bare `Enter`.
4. Herefter starter OCR-processen automatisk:
   - Alle billedfiler (*.tif *.tiff *.jpg *.jpeg *.gif) i den valgte mappe og dens undermapper bliver behandlet.
   - Mapper der **starter med `-`** bliver **sprunget over**. (f.eks. mappen -originalfiler og -scan bliver sprunget over)
   - Filer, hvor der **allerede findes en tilhørende OCR-PDF**, springes over.
5. Når det er færdigt, vises en besked — eller pc'en slukkes automatisk, hvis du valgte det.
6. Resultatet bliver gemt som PDF i **samme mappe som billedet**, med **samme navn** (f.eks. `filnavn.pdf`).

Logfil

Alle handlinger bliver logget i filen:
`ocr_log.txt`
Denne logfil ligger i samme rod-mappe, som dine behandlede filer.

Findes logfil i forvejen, tilføjes data blot til logfilen.