InfoShare PDFtoText Scanner
Local Windows version konfigurationsfil

Overblik:
PTS programmet behandler indscannede pdf filer I de sourcemapper, der er angivet I ptsconfig.ini.
- ptsconfig.ini kan redigeres med notepad eller programmet PTSconfig.
Selve dokumentbehandlingen er beskrevet I destinationsmappens zones.txt (eller rules.txt) fil.
- zones.txt og rules.txt kan redigeres med notepad eller programmet PTStools.

Installation af software
PTS softwaren installeres på en Windows PC v.h.a. installationsscriptet scanmonitor_install.exe. Installationen starter automatisk programmet PTSConfig, som opretter en standard ptsconfig.ini fil i PC'ens %appdata% mappe. Denne konfigurationsfil tilhører altså den enkelte brugerprofil (man kan have forskellige profiler for forskellige brugere på PC'en). Konfigurationsfilen kan efterfølgende tilrettes med en almindelig teksteditor eller ved at køre PTSConfig programmet igen.

Konfigurationsfilen indeholder en eller flere blokke med beskrivelse af source-mapper. For hver blok er angivet en række opsætningsparametre, som gælder for pdf-filer i netop denne mappe. Kommentarerne nedenfor skal ikke findes i den konkrete konfigurationsfil; de udfyldte parametervalg er eksempler.:

[PDFScan1]
Navn på denne konfigurationssektion - der kan være flere sektioner og navn kan frit vælges.
source = n:\scan\indmappe1
Her skal programmet se efter indscannede dokumenter (i pdf-format). En lokal Windows-mappe eller en delt netværksmappe.
Hvis parameteren er udeladt: Hele profilen overspringes
destination = n:\scan\udmappe1
Her skal programmet lægge det behandlede dokument (som søgbar pdf).
Hvis parameteren er udeladt: Hele profilen overspringes
setsize = 0
Antal sider i den indlæste pdf fil, der søgemæssigt skal betragtes som et enkelt dokument. Scanneren afleverer en enkelt pdf-fil, men vi kan med denne parameter vælge at betragte det som et bundt af blanketsæt med samme sideantal. Gyldige værdier: Heltal (men da scannerens indbakke har en begrænset kapacitet vil større tal ikke give mening. Bemærk at idet de to sider af et indscannet dokument betragtes som to separate dokumenter, så vil en dobbeltsidig indskanning af enkelte dokumenter kræve at setsize sættes til 2. Værdien 0 betyder ingen setsize (alle sider i én scanning udgør ét dokument)
Hvis parameteren er udeladt: setsize er 0
documentsize = 1
Den afleverede pdf fil kan bestå af flere set. Parameteren angiver hvor mange sæt, der indgår i den afleverede pdf fil. Ekempel: 10 indscannede sider er 5 blanketsæt bestående af to forskellige blankettyper, som kan ligge i vilkårlig rækkefølge: setsize=2 og documentsize=2 vil producere færdige pdffiler med 4 sider, hvor alle sider er blevet genkendt efter reglerne for side 1 og 2.
Hvis parameteren er udeladt: documentsize er 1
savetext = 0
0=Gem ikke tekstindholdet af filen, 1= gem tekstindholdet sammen med pdf-filen. Tekstudgaven indeholder den OCR-fortolkede tekst fra dokumentet. Den kan være nyttig i en testsituation.
Hvis parameteren er udeladt: savetext er 0
fullocr = 0
0 = Der kigges kun på indholdet af zoner og det fulde dokument OCR-fortolkes ikke. Den resulterende pdf er ikke søgbar, men behandlingen er hurtigere. Indstillingen giver ikke mening,

hvis indhold behandles med en rules fil – kun hvis der er tale om en zones fil. 1 = der laves OCR fortolkning på hele dokumentet.
Hvis parameteren er udeladt: fullocr er 0
singlepdf = 0
1=Alle sider fra det indlæste scanjob samles, uanset setsize, til én pdf fil, 0= Der dannes en pdf fil for hvert sæt.
Hvis parameteren er udeladt: singlepdf er 0
backup = n:\scan\backup1
Læg ucensureret version (udgave uden sværtning af områder) af de indscannede dokumenter i den givne mappe. Dvs. at pdf filer i denne mappe ikke er blevet anonymiseret selv om den tilhørende zones.txt fil indeholder anonymiseringsinstrukser.
Hvis parameteren er udeladt: backup er slået fra
landscape = 0
0=indscannede dokumenter er vertikale (default), 1=indscannede dokumenter er horisontale.
Hvis parameteren er udeladt: landscape er 0
autodeletebackup = 0
0 (eller ’no’ eller ’NO’ eller ’false’) = Der foretages ikke automatisk sletning af pdf filer i backup mappen. 1 = Pdf filer i backup mappen slettes automatisk når den tilhørende pdf fil i destinationsmappen slettes.
Hvis parameteren er udeladt: autodeletebackup er 0
raw_backup = n:\scan\backupraw
Med denne parameter udfyldt vil der i den angivne mappe blive lagt en pdf fil med samtlige indscannede sider, uanset setsize og documentsize.
Hvis parameteren blank eller udeladt: Der laves ikke nogen backup af det samlede scanjob.
blank_separation = 1
1 = setsize ignoreres og der separeres efter blanke sider (% mean threshold er sat til 99.95% - der må altså godt være en lille smule sort på papiret). 0 = adskillelse ved blanke sider er ikke aktiv.
Hvis parameteren er udeladt: blank_separation er 0
no_sep_at_blank_back = 1
1 = forhindrer at blank bagside af papir tolkes som separator (ved automatisk scanning af begge sider af et ark), 0 = Alle blanke sider, også bagsider, opfattes som adskillelsesside.
Hvis parameteren er udeladt: no_sep_at_blank_back er 0
rotate = page-no,x,y,width,height####regex
Hvis parameteren findes i konfigurationsfilen vil det automatisk blive undersøgt om dokumentet er blevet scannet med nederste kant først. Undersøgelsen består i søgning efter en bestemt tekst i et bestemt område, fx rotate = 1,200,200,400,150####CPR-nummer. Bemærk at dette øger behandlingstiden.
Hvis parameteren er udeladt eller uden tildelt værdi: rotate er slået fra
double_rotate_check = 0
1: tjekker igen efter siden er roteret om der er et match, hvis ikke der er et match ignoreres roteringen, 0: Ingen dobbelttjek – hvis rotate indstillingen ikke har fundet den eftersøgte tekst bliver siden roteret og anvendes efterfølgende i sin roterede form.
Hvis parameteren er udeladt: double_rotate_check er 0
counter_file = N:\scan\taeller.txt
Denne parameter angiver filnavnet, der indeholder en tæller (løbenummer) for de indscannede dokumenter. Hvis flere profiler peger på den samme tællerfil vil de forskellige profiler benytte et fælles opdateret løbenummer; hvis hver profil har sin egen tællerfil vil løbenummeret være fortløbende indenfor dokumenter tilhørende den pågældende profil.
Hvis parameteren er udeladt: counter_file er ”[source-mappe]/loebenr.txt”

psm=4
Page Segmentation Method er en detaljeret styring af Tesseracts fortolkning af sidens tekst.
Tesseract kender følgende metoder:

0 Orientation and script detection (OSD) only.
1 Automatic page segmentation with OSD.
2 Automatic page segmentation, but no OSD, or OCR.
3 Fully automatic page segmentation, but no OSD. (Default)
4 Assume a single column of text of variable sizes.
5 Assume a single uniform block of vertically aligned text.
6 Assume a single uniform block of text.
7 Treat the image as a single text line.
8 Treat the image as a single word.
9 Treat the image as a single word in a circle.
10 Treat the image as a single character.
11 Sparse text. Find as much text as possible in no particular order.
12 Sparse text with OSD.
13 Raw line. Treat the image as a single text line, bypassing hacks that are Tesseract-specific.

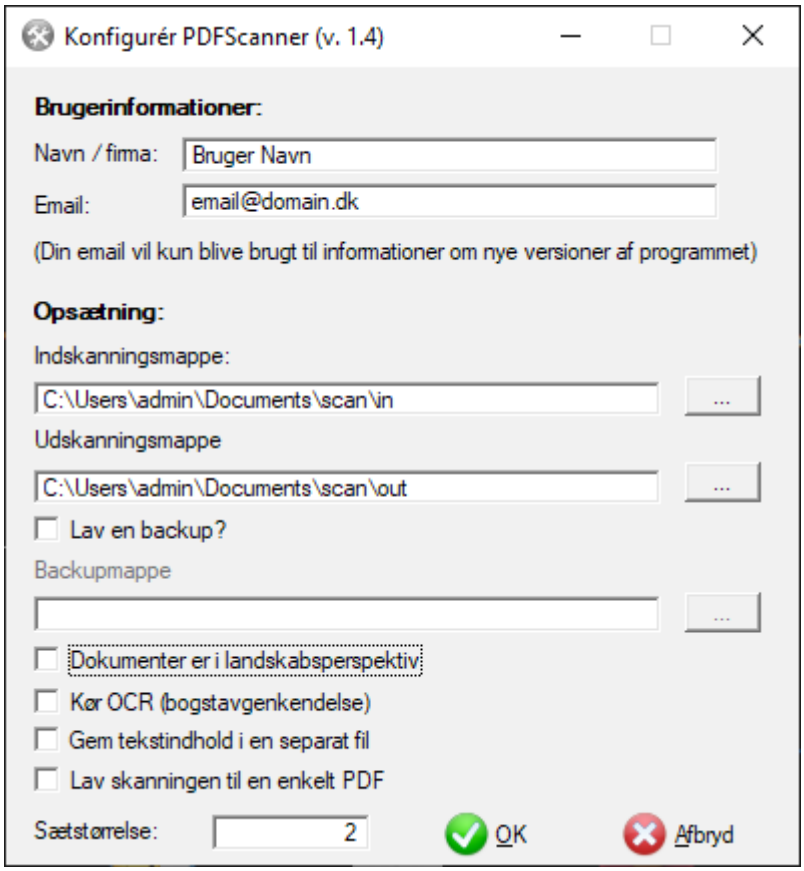
Hvis parameteren er udeladt: psm er 4

Nogle indstillinger kan være logisk i modstrid med hinanden eller give uventet funktionalitet.
Hvis AutoDeleteBackup er slået til og konfigurationen indeholder flere profiler, skal disse profiler enten benytte samme tællerfil eller have hver sin backup mappe. (Autodelete fungerer ved at overvågningsprogrammet kontrollerer hvilke filer, der findes i backupmappen, men ikke i destinationmappen – sådanne filer slettes.
Såfremt destinationmappen ændres til en absolut sti ud fra kommandoer i rules eller zones vil der ikke blive taget backup.

Enhver linje i ptsconfig, der starter med **#** eller **;** opfattes som en kommentar.

Vejledning til PTSConfig programmet

Installationsproceduren installerer programmet PTSconfig, der dels benyttes til at registrere ejernavn til brug for licensgodkendelse og dels er en hjælp til oprettelse af en standard konfigurationsfil med de almindeligste parametre.



Ved brug af programmet skal de første 4 felter udfyldes. Klik på OK gemmer dine indstillinger i filen ptsconfig.ini i den Appdata mappe (skriv %appdata% i adressen i stifinder), som hører til din Windowskonto (I undermappen infoshare/PTS/). Ekstra brugere på samme PC, hvor PTS programmerne allerede er installeret, kan oprette deres egen konfigurationsfil uden at skulle installere programmet igen ved blot at køre PTSconfig. Af hensyn til automatisk licensgodkendelse skal ejerfeltet udfyldes med samme brugernavn.

Eksempel på ptsconfig.ini fil

```
[1ark]
source = J:\scan\1ark
destination = J:\scan\behandlede\1ark
setsize = 2
singlepdf = 0
savetext = 0
fullocr = 0
backup = J:\scan\backup
[2ark]
source = J:\scan\2ark
destination = J:\scan\behandlede\2ark
setsize = 4
singlepdf = 0
savetext = 0
fullocr = 0
backup = J:\scan\backup
[3ark]
source = J:\scan\3ark
destination = J:\scan\behandlede\3ark
setsize = 6
singlepdf = 0
savetext = 0
fullocr = 0
backup = J:\scan\backup
```

InfoShare
December 2019