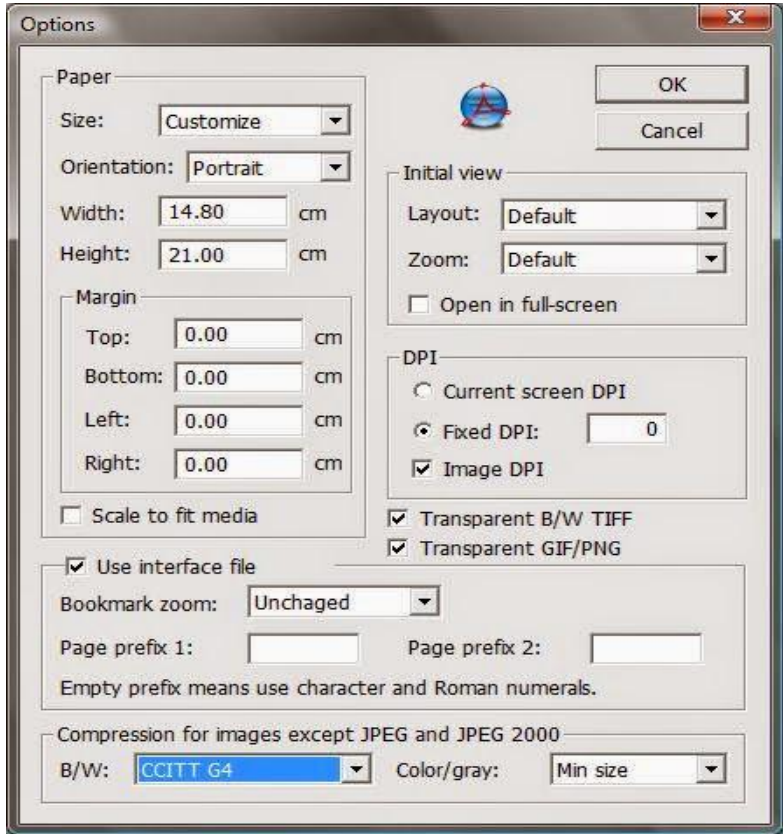# GENERATE PDF

The software differs a little depending on OS. But compression is very important:

- for 1-bit binary images (Black and White in ScanTailor) always use CCITT-G4 compression.
- for the other types, use ZIP/Deflate compression. If the detail is not important (e.g. cover pages) you can use JPEG compression with 75-80% quality (that is lossy).
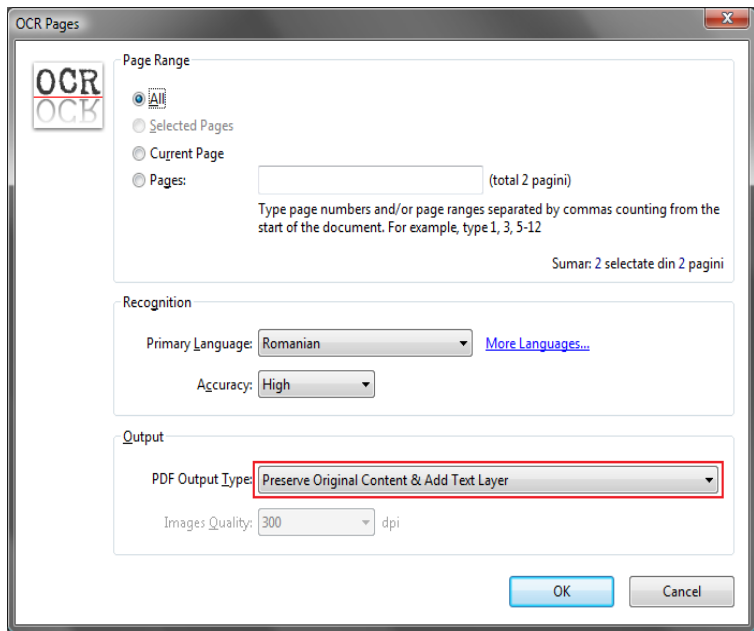
Windows

You have a bunch of TIFFs that need to be joined into a single PDF. You can use the utility FreePic2Pdf. Add the folder / files then click on **Options** button. Set the paper size greater than image size, otherwise images will be cropped. This software determines compression based on color depth of each image. If you look at the Compression options, it says JPEG and JPEG2000 are excluded. That is because those formats are embedded as-is in the PDF file without any other conversion.

Once you have a PDF file, you may want to make it searchable. So download PDF-XChange Viewer and its OCR language pack and open the PDF with it. Click the **OCR** button (or **Document** menu - **OCR pages**), select language and start it.

FreePic2Pdf options dialog

PDF-XChange Viewer OCR dialog

Linux

Things are more difficult on Linux and there is a lack of GUI tools. FreePic2Pdf runs good on Wine but PDF-XChange Viewer doesn't.

If you want OCR, Tesseract can be used. Starting with version 3.03 (which is still beta at the time of writing this) it supports PDF output. But it has a problem with compression on color images. PDFs resulted from it are larger than the ones from FreePic2Pdf and PDF-XChange OCR. Let's write a script that will run Tesseract. First of all install (replace<lang> with a three-letter language code):

```
sudo apt-get install pdftk tesseract-ocr tesseract-ocr-<lang>
```

This is the script (add an extra "f" at *.tif if your files have .tiff extension):

```
#!/bin/bash
LANG=eng #replace with your language code
shopt -s nullglob
for f in *.tif; do
 echo "Running OCR on $f"
 tesseract -psm 1 -l $LANG $f $f pdf
done
echo "Joining files into single PDF..."
pdftk *.pdf cat output ../outdocument.pdf
rm -r -f *.pdf
```

Save it as ocr.sh, make it executable (chmod +x ocr.sh) and run it in the same

directory with the TIFF files. The output result will be one folder up.

If you don't want OCR, all you have to do is convert all TIFF files to PDF and join

them. If you're thinking to use ImageMagick - it is quite slow. More than this, the

color depth must be specified for each image. That's ok, but if the document is

made of different color depth pages the script gets complicated. By far the fastest

solution, Leptonicalibrary supports PDF joining too. But there's no utility to do

that. You'll have to compile a small program. Install the build tools and

dependencies:

```
sudo apt-get install build-essential gcc liblept4 libleptonica-dev
```

Now save the following code as joinpdf.c

```c
#include <leptonica/allheaders.h>

#include <stdio.h>

int main(int argc, char *argv[]){

  if (argc < 3) {

    printf("Not enough arguments specified!\njoinpdf <input_folder> <output_pdf>\n");

    exit(1);

    }

  if (argc > 3) printf ("Too many arguments!\nOnly the first two will be taken into account.\n");

  int r = convertFilesToPdf(argv[1], NULL, 0, 1, 0, 0, NULL, argv[2]);
```

```
   if (r == 0) printf ("%s successfully written!\n", argv[2]);

   else printf("Conversion failed");

  }
```

In the same folder with the joinpdf.c file run in a terminal:

```
gcc -o joinpdf joinpdf.c -llept
```

To use it, you can run it from the directory you are in (./joinpdf) or you can copy it system-wide (sudo cp ./joinpdf /usr/bin/joinpdf). Two arguments must be passed: the folder that contains TIFF files and the output PDF file with extension (e.g. joinpdf ~/Documents/Scan/out ~/Desktop/mydocument.pdf). This little app is based on Leptonica and automatically detects color depth of each image and adjusts the compression type.

If you find this too difficult, you could run FreePic2Pdf on Linux with Wine.

OS Differences

The Linux tools do not alter the page size when converting to PDF. On Windows, page size can be set in FreePic2Pdf.

Source: http://onetransistor.blogspot.in/2014/11/high-quality-scanning-and-low-file-size.html