

## Creating an Text-Searchable PDF from individual Image Files

Programs used:           **Photoshop CS4**  
                              **Adobe Acrobat 9.0 Pro**

Notes:

- This process works with TIFF and JPEG at least (we didn't have an occasion to try other file formats). JPEG2000s had to be converted into TIFF format before any additional steps could be taken (Step 1).
- This workflow also includes instructions to batch resize images, due to size restrictions in Adobe Acrobat Pro 9.0's OCR processing ability (Step 1A). For text recognition to work in Acrobat Pro 9, the longest side of any page in a PDF must be shorter than 45 inches.

If you are already working with TIFFs or JPEGs that are less than 45 inches along the longest side, skip to Step 2.

### 1       **Converting JPEG2000s into TIFFs or JPEGs**

- In Photoshop, select **File →Scripts →Image Processor**.
- In the following dialog box, select the folder where the JPEG2000 files are located.
- Select **Save in Same Location** (Photoshop will create a subfolder, so the original files will not be overwritten).
- Select TIFF or JPEG as the file type.
- Click **Run** to begin convert.

### 1A      **Converting JPEG2000s into TIFFs or JPEGs and/or Reducing TIFF/JPEG dimensions**

- In Photoshop, select **File →Scripts →Image Processor**.
- In the following dialog box, select the folder where the image files are located.
- Select **Save in Same Location** (Photoshop will create a subfolder, so the original files will not be overwritten).
- Select TIFF or JPEG as the file type, click the **Resize to fit** box, and set your maximum pixel length value(px) for both width and height. Photoshop will scale according to dimensions, so there will be no warping. (You can check the dimensions of your image by selecting **Image →Image Size...**)
- Click **Run** to begin resizing.
- Complete walkthrough available here: <http://www.digital-photography-school.com/how-to-batch-resize-in-photoshop>

## 2 Creating PDFs and Making Them Text-Searchable in Adobe Acrobat 9.0 Pro

- In Adobe Acrobat Pro, select **Create** from the top left corner (directly below the File, Edit, and View dropdowns).
- Select **Merge Files into a Single PDF**
- Click **Add files...** and locate the folder of image TIFFs or JPEGs. (You may need to switch the **Files of Type** dropdown at the bottom of the search box to **All files** to locate the folder containing the TIFF files.)
- Select the files that need to go into the PDF, then use the **Move Up/Down** buttons or drag files into the correct order.
- Click **Combine Files** to create the PDF.
- A **Save As** box should come up. Save the item.
- Select **Document** → **OCR Text Recognition** → **Recognize Text Using OCR**, then select **All Pages** to run the character recognition thingamabob.

## 3 Checking/Editing OCR for Typos, Unrecognized text (Optional)

*This is definitely a time consuming and optional step. Acrobat Pro's character recognizer is of acceptable quality, but may have trouble recognizing some words, especially from older texts and typefaces. This may be helpful as a workaround, but we found it to be a little too difficult to bother with, instead opting to upgrade our OCR program. ABBYY FineReader came highly recommended from Internet Archive.*

Select **Tools** → **Advanced Editing** → **Touch Up Text Tool** to begin editing the OCR.

- 1) With the Text TouchUp tool, select all text on the page (ctrl-A).
- 2) Right-click the page and click "Properties".
- 3) Go to the "Text" tab and select a Fill color for the font. Now the overlaid text is visible!
- 4) Make any text corrections needed.
- 5) Select all text again and choose "No Color" for the fill text to make it invisible again.
- 6) Save.

(These instructions came from a user forum on Adobe for Acrobat Pro 8, but worked with Acrobat Pro 9.)

*Prepared by Mike Shea and Cathy Billings of the Glendale Public Library as a part of their 2011-12 LHDRP grant.*