

Creating Searchable PDFs with Adobe Acrobat XI - Quick Start Guide

Overview

Many scanned PDF files are not usable with technology because the content is perceived as images by computers and other devices. These files need to be fixed so that devices can recognize the text. To do this, we need software that can process files using **OCR (Optical Character Recognition)**.

OCR is software that recognizes and interprets text in an image and converts it to text that a computer can read.

Adobe Acrobat Professional (the type that can create PDFs, not just read them) is one of the cheaper and more common methods available with OCR, and is the focus of this Quick Start Guide.

Running OCR on a file has many benefits, including:

- Searchable text
- Copy and paste capability
- Text-to-speech capability
- Screen reader compatibility

Creating a Searchable PDF

The minimum you should do to make a PDF accessible is to create a **Searchable PDF** using OCR.

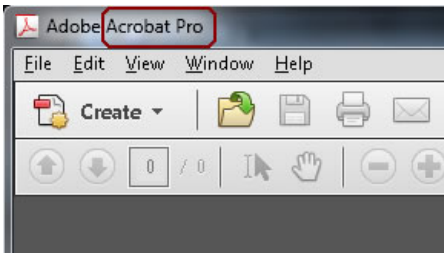
OCR at the Scanner

Some scanners and photocopiers have the ability to OCR documents as they scan so that you don't have to fix it afterwards. Many Xerox photocopiers can scan to PDF via email or network, and you can select the option to make that PDF searchable.

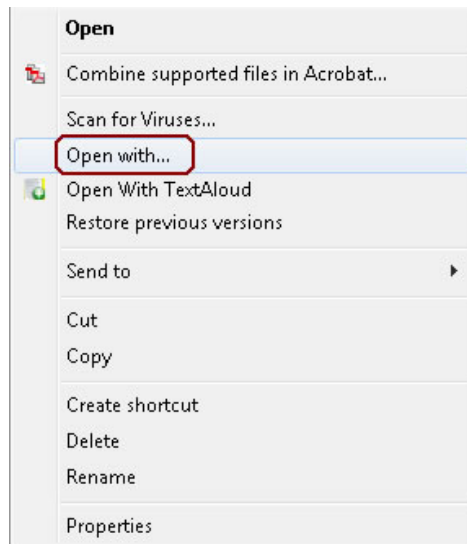
OCR with Software

In case you are unable to OCR at the source, it is easy to make a document searchable using software such as Adobe Acrobat Pro. This guide focuses on using Acrobat XI, but the process is similar in other versions.

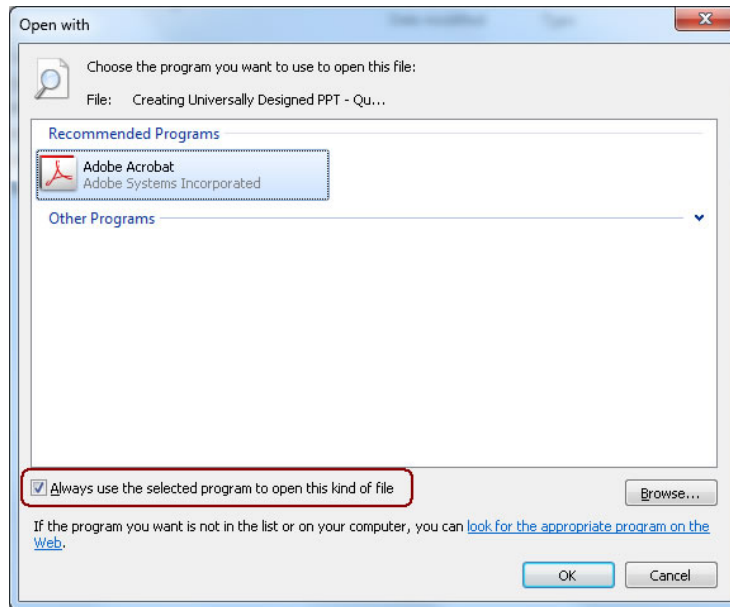
- Open your PDF in Acrobat Professional. (Make sure you are not in Adobe Reader, as you cannot edit a PDF in the Reader version.)
- To see if you are in the correct program, check the top left of your document window. It should say Adobe Acrobat Pro:



- If you are in Acrobat Pro, skip to the next section, *Use the "Text Recognition" Toolbar*.
- If your document says Adobe Reader, close the program.
- **Right-click** on the document and select **Open With:**

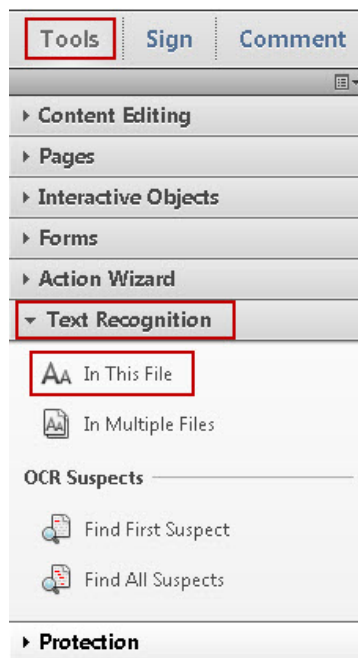


- Choose **Adobe Acrobat**. The choices on your computer may look different than the screenshot below.
- To make Acrobat your default program for PDFs so you don't have to repeat this process, make sure the box is checked to **"Always use the selected program to open this kind of file."**

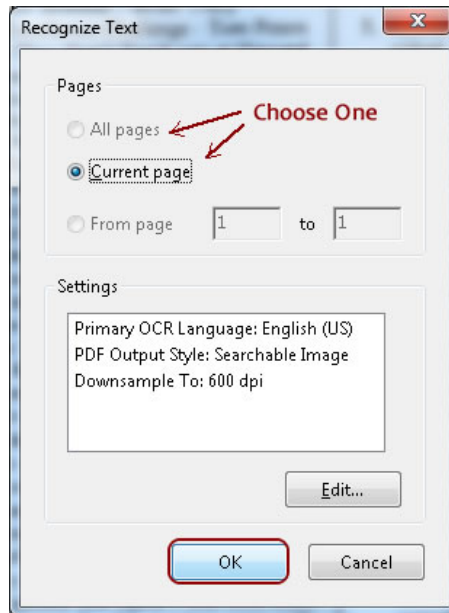


Use the “Text Recognition” Tool

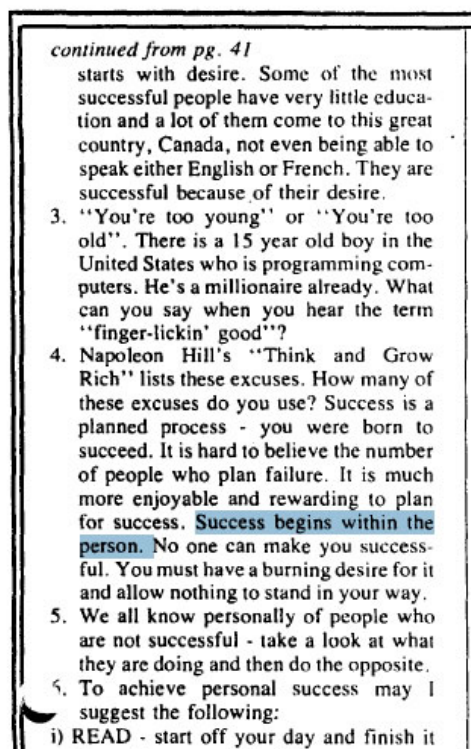
1. Now that your document is open in the correct software, click on **Tools** to bring up the Tools Pane on the right side of the screen.
2. Click on **Text Recognition** to expand the menu and select **In this File**.



3. When the **Recognize Text** box comes up, choose how many pages to OCR at once. If it is a large file, try only a few at first, since it could take a while. Otherwise select the **All Pages** radio button and click **OK**.



4. Once the process finishes running, you should be able to select text in your PDF file and you now have a searchable PDF. Simply save the document and it is OCR'd forever.

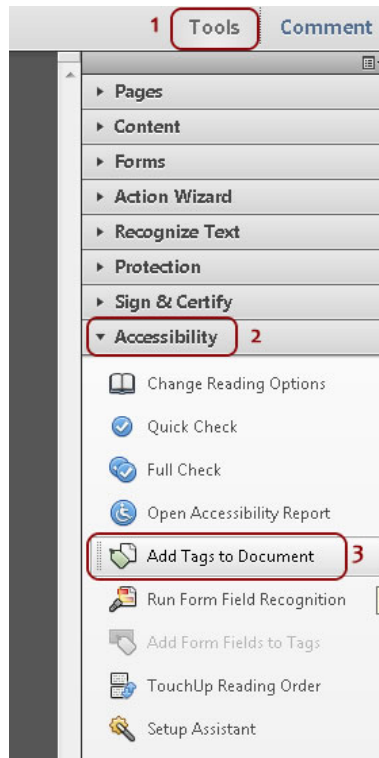


Adding Document Structure

Add Tags to the Document

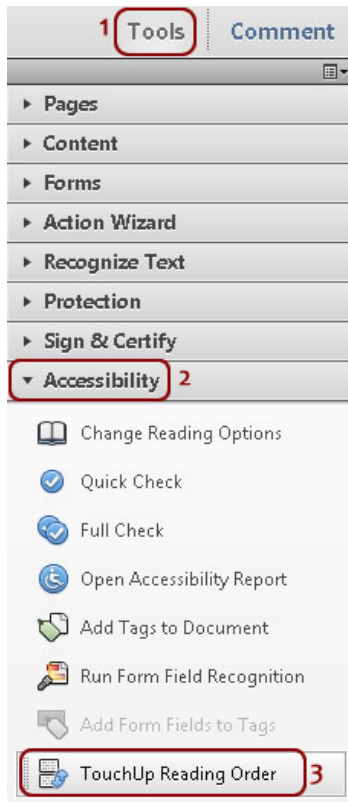
Many users listen to documents rather than reading them. To double-check that your document will be read in the correct order, you need to add Tags. Tags provide document structure, which creates a table of contents that is useful for pre-reading, scanning, and navigation.

1. To add tags, open the **Accessibility Tool**.
2. In the expanded Accessibility tools menu, select **Add Tags to Document**.

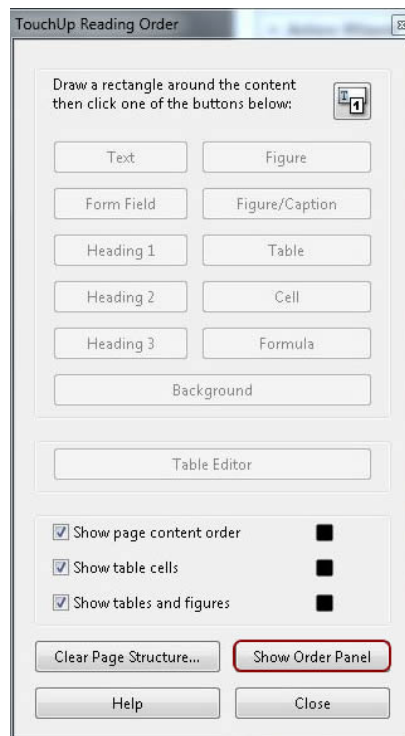


Check the Reading Order

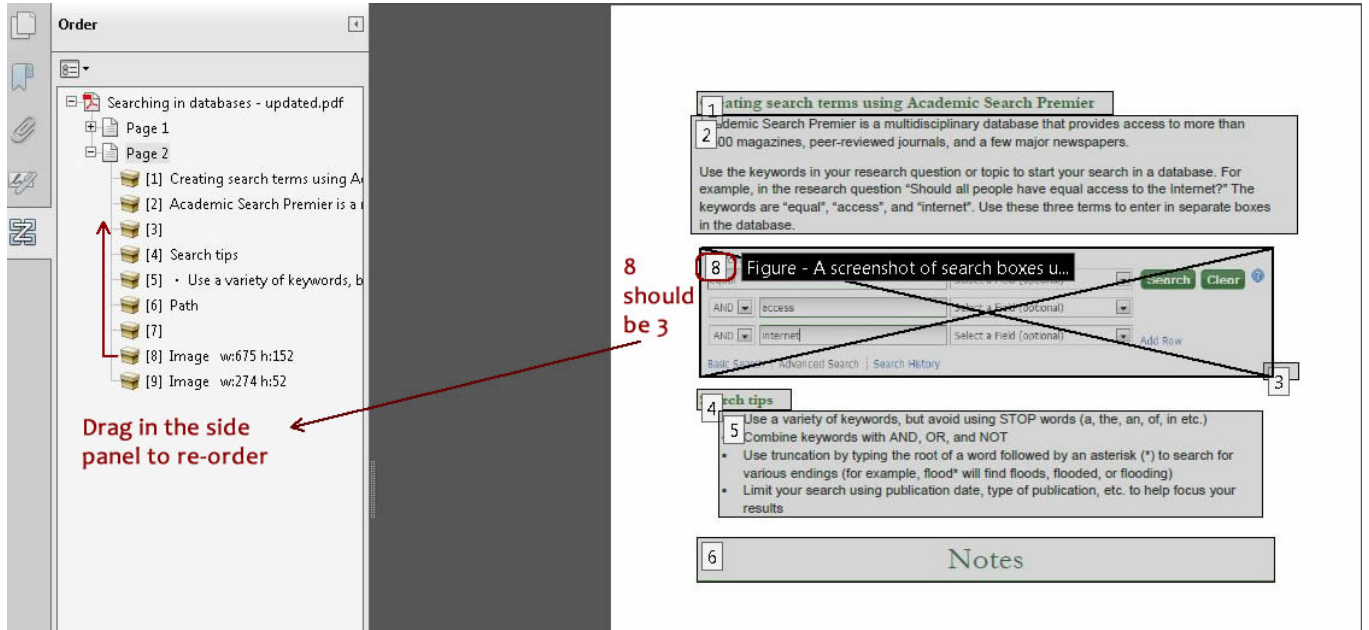
1. In the Accessibility Toolbar, click on **TouchUp Reading Order**.



2. In the new window that pops up, click on **Show Order Panel**.



- Your page should fill with sections that are numbered. If they are not in the correct order, **Select** and **Drag** the numbers into the right order using the pane on the left side of the screen.



- Check each page in the document and save.

Other Accessibility Considerations

Although not covered in this guide, keep in mind that in order to make the PDF fully accessible, images, graphs and charts need alternate text, and tables need to be configured with header rows. For now, making the PDF searchable and adding tags are the most important aspects to incorporate in your PDF workflow.