# Step 1. Preparing a Word Document from a Text-selectable PDF

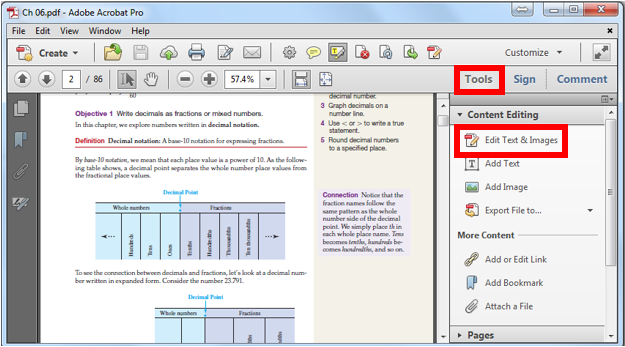
The first step is to take a text-selectable PDF and “clean it.” This is the process of rearranging content and deleting images, background colors, tables, etc. Cleaning not only helps INFTY run faster, but it makes the document easier to edit.

Before starting any of the steps below:

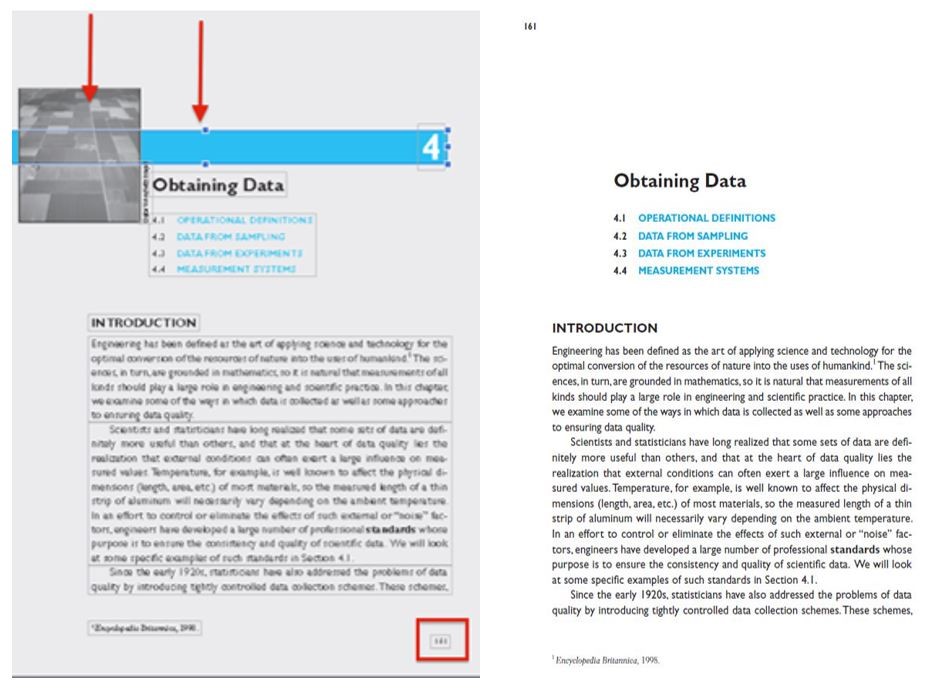
* The pages from the source PDF must be extracted.
* Extracted file must be saved separately from the source.

## Steps

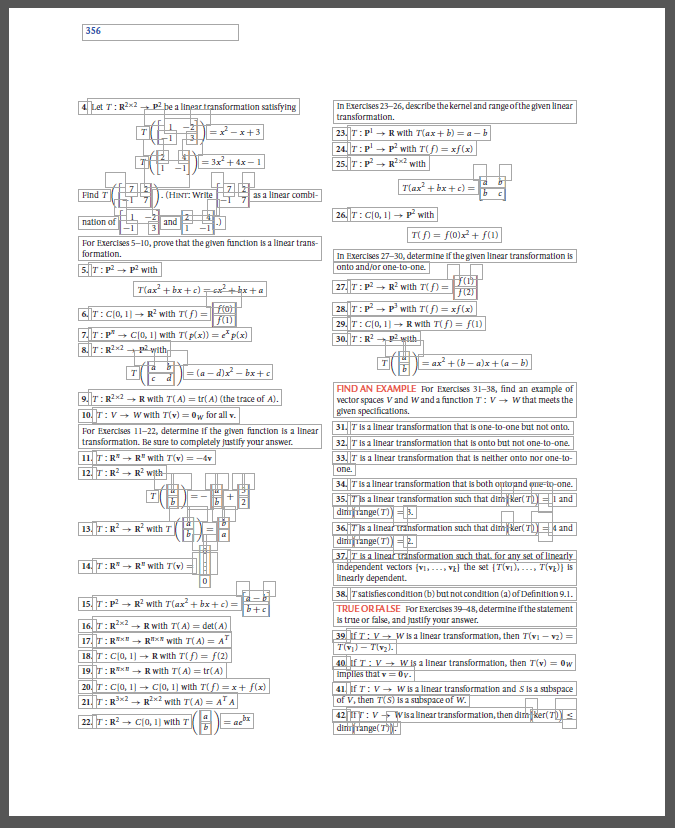
1. Open the PDF.
2. Select the **Tools** option on the right hand.
3. Select **Edit Text & Images** as shown in the image below. Notice that borders appear around text and images.



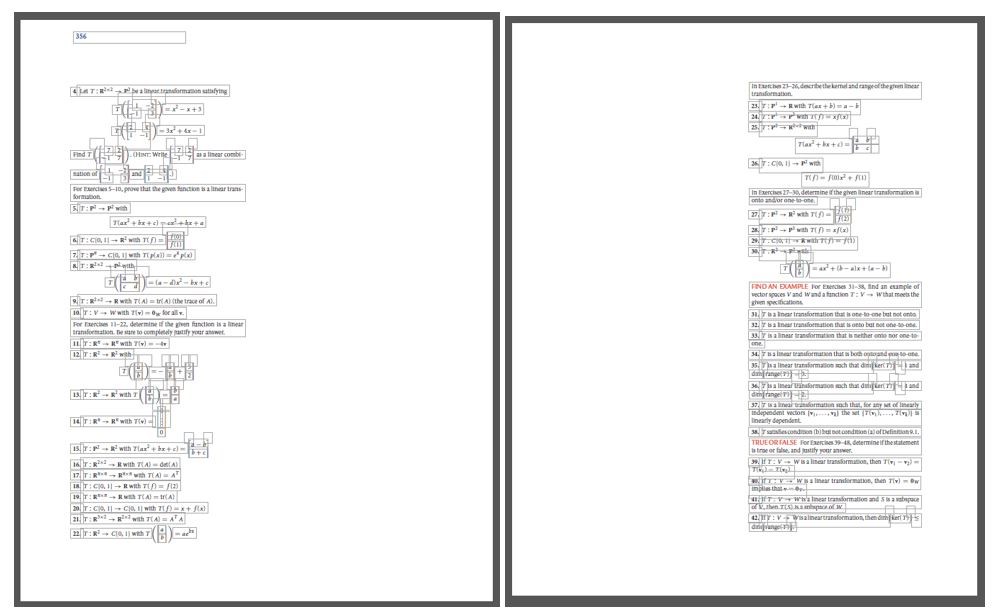
1. Delete the following: (To remove the contents of these borders, right-click and select **Delete** from the resultant menu or select the border and press the **Delete** key.)
   1. Lines/line art.
   2. Images.
   3. Background.
   4. Recurring page headings.
   5. Tables.
   6. Borders.
   7. Shading.
   8. Other superfluous decorative features (i.e., "eye candy").



1. If something is too time-consuming to remove (like a colored background that refuses to be deleted), leave it and correct it late when editing the Word document.
2. **INFTY** typically makes mistakes with images and when columns of text appear next to each other. When there are multiple columns of texts on a single page (often found in the exercises section towards the end of a section), add blank pages with the keyboard shortcut **CTRL+SHIFT+T**.
   1. BEFORE: Two columns of text on a single page.



* 1. AFTER: Two columns of text on two separate pages.



1. Make sure page numbers are at the **top left** of each page. Failure to do so will cause problems when OCR'ing in INFTY. **INFTY reads from top to bottom, left to right**, so ensure that all text is in a singular column in the correct order.
2. When the PDF has been thoroughly cleaned, save the PDF as TIFF images (in a separate folder for every chapter/section). Go to **File > Save as other > Image > TIFF**.
   1. Save these in one folder on your desktop for convenience in the next step. As this folder will be deleted later, naming is not important. Ensure the follow settings are selected:

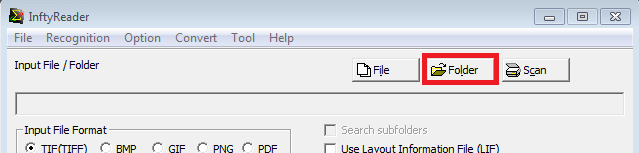
Save as TIFF settings
Monochrome LZW
Grayscale None
Color None
RGB: Off
CMYK: Off
Grayscale: Off
Colorspace Monochrome
Resoultion 600 pixels/inch

# Step 2: INFTY

* 1. Open up **INFTY Reader**, not **INFTY Layout**.
  2. Make sure these settings are checked when running**INFTY**:

InftyReader screenshot
Select Tiff, English, 600 dip, All math symbols, Microsoft word (xml), the end of each paragraph, no to open result file by the related application

1. Select **Folder** and select desired folder of TIFF images (the one you made at the end of the cleaning process) then select **Start**.



1. Once finished, an .xml file will appear in the TIFF folder.
2. Save as a Word document (.docx) file.
3. Trash the folder of TIFF images, and the log created by **INFTY**. The only thing you need now is the *.docx* document.

# Troubleshooting INFTY

If **INFTY** isn't giving an .xml file (i.e., giving a KML file or just a log):

1. Check the PDF saving settings to make sure they were set as TIFF images. [Settings.]
2. Try splitting TIFF images into smaller batches for**INFTY** to process. See if processing only 10 – 20 images at a time works. [Less TIFF images]
3. If the above steps don’t work, create a folder within the TIFFs folder, and move the images there. Have**INFTY** process this nested folder. [Longer path name]
4. Try to have **INFTY** process the folder of TIFFs again, a few more times.
5. If nothing has worked, process the PDF through **ABBYY** and manually enter in equations. However, this situation is **rare**..
6. If the output from I**NFTY** is a .kml file instead of .xml, avoid making the output path onto a network drive. Save onto the computer’s hard disk, then move it onto a network drive.