

Scanning Text Documents Using Optical Character Recognition (OCR)

Unlike scanning a document as an image, which renders the text uneditable, scanning a text document using optical character recognition (OCR) software makes it possible to edit the scanned document in a Word Processing or text editor program. The only OCR-capable scanner at CURA is the HP 1100A printer in Peggy Wolfe's office. To use the OCR capability of the printer, you must do the following:

1. Turn on Peggy's computer by turning on the power strip and then pushing the start button on the front of the machine.
2. Open the OCR software program. Choose "Start" from the Windows menu, then select "Programs," "ReadIris OCR," and then "ReadIris."
3. Remove any staples, creases, etc. from the document you wish to scan.
4. Place the document in the slot at the front of the printer. The page should be placed in such a way that the printed side of the document is toward you and the top of the page is head-first in the slot. Be sure to adjust the paper guide so the document is straight when it feeds through the scanner.
5. To begin scanning, simply press the "Auto" button. This uses standard settings that will work for most uses. If you wish to learn more about specialized settings, use the "Help" menu.
6. When the scan is complete, you will be prompted to save your document as an RTF file. Make sure you save the document to a floppy disk, the desktop, or somewhere else on the computer where you can find it. The resulting RTF document can be opened in any standard word processing program.
7. Once you have saved the file, you will be prompted to scan another page or end the session. Choose "yes" if you want to scan another document or "End" if you're finished. NOTE: If the document you are scanning is two-sided, you must turn the document over and repeat the process above to scan the second side.

8. Shut down the computer once you are finished by closing all programs, choosing "Start" and then "Shut Down," and turning off the powerstrip once the shutdown process has been completed.

9. Cleaning up your document. Although the scanner does a pretty good job, it does have some limitations. It cannot distinguish footnotes from regular text and it cannot separate headers and footers from the body of the document. You will need to reformat the document in a word processing program if you want to retain these elements. Formatting (bold, italic, etc.) and font type and size will likely be different. However, it can recognize a wide variety of different fonts, font sizes, etc., and it sure beats retyping the entire document. The resulting RTF document can be reformatted to use the original font type (if you know what it is). It is especially important to SPELLCHECK scanned documents, as the scanner can misinterpret characters (that is, it might read a lowercase letter l as the number 1, or a zero as the letter O).