

# What is OCR?

OCR stands for "Optical Character Recognition." It is a technology that recognizes text within a digital image. It is commonly used to recognize text in scanned documents and images.

OCR software can be used to convert a physical paper document, or an image into an accessible electronic version with text. For example, if you scan a paper document or photograph with a printer, the printer will most likely create a file with a digital image in it. The file could be a JPG/TIFF or PDF, but the new electronic file may still be only an image of the original document. You can then load this scanned electronic document it created, which contains the image, into an OCR program. The OCR program which will recognize the text and convert the document to an editable text file.

## How Does OCR Work?

OCR software processes a digital image by locating and recognizing characters, such as letters, numbers, and symbols. Some OCR software will simply export the text, while other programs can convert the characters to editable text directly in the image. Advanced OCR software can export the size and formatting of the text as well as the layout of the text found on a page.

## Does OCR Create an Accessible Document?

The short answer is no, not really. Some OCR programs allow you to scan a document and convert it to a word processing document in a single step, but this still may not be an accessible document. Once you use the OCR to process your document, you must select the text and read it to verify the process was successful, and the text makes sense. You may have to spell check it, add headings, add tags, reorder it and more. You can do this with your word processor such as Word, or Adobe Acrobat Pro.

## Do I really need to Proofread and Correct an OCR output?

Yes! Think of it this way: If your original had really good contrast and readability a 99% success rate is possible with some OCR software, but what if the 1 % wrong was the tuition rate for the college? If the original image had poor contrast and readability the success rate could go down to 50% or even be unreadable. You won't know until you check it!

## What if My OCR Output is Really Bad?

If your original document or image has poor contrast, fuzzy characters, overlaps, etc. the OCR software may recognize the text, but the text may not be accurate and be hard to read to make your OCR output is more successful with easier to verify and correct output, make sure your original is not a fuzzy reprint. The original should have good contrast and sharp letters. If you cannot get your hands on a better version of the original, the printer, or printer software, may have settings on it that will produce a better scan.

## Where Do I find OCR software?

OCR conversion software may be installed on your printer or the software it came with. If you need help finding out if your printer has it, or setting it up, contact the NECC Help Desk.

Adobe Acrobat Pro has an OCR conversion included with the software.

## References

- [Explain that Stuff: Optical Character Recognition](#)
- [Tech Terms Software Terms: OCR Definition](#)