



The Pros and Cons of Optical Character Recognition

June 21, 2021



Throughout your journey of becoming a [paperless](#) company, at some point you may come across the term [Optical Character Recognition](#) (OCR.) OCR refers to the process by which scanned images are electronically “read” to convert them into editable text. This allows for users to edit the digital document and will allow you to modify the content as you see fit. Sounds great, right? It is! However, when it comes to OCR, we have seen a lot of misconceptions about the pros and cons of the software, especially when used within a document management system.

OCR in a document management system is usually used for [indexing features](#). At first glance OCR sounds like the easiest and fastest way to allow for your scanned documents to be editable and searchable. There is some truth to that statement, but overall there are some initial benefits and drawbacks to consider before implementing it within your document lifecycle. Through this blog we will go over the major pros and cons of OCR and the most useful ways to use it within a document management system.

Con: Accuracy and Reliability

OCR accuracy is never 100%. As we said previously, OCR is a technology that examines scanned images and attempts to read the text that it finds. The technology does this by going through every single letter of the document and trying to compare the pixel it sees in the image with pixels for the letter A, B, C and so on. Although it usually does reasonably well on printed documents, with so many different fonts and similar letters, it is possible the OCR technology could make errors. For example, it may misread letters, mix together

text, or skip over unreadable letters. Due to the high risk of errors, it is best practice to have a clean-up of the electronic text after it goes through the OCR process to check for any mistakes.

Recognition of handwritten documents with OCR is also extremely limited. Each person has very unique handwriting which can make it almost impossible for the OCR technology to get every single letter correct. With the evolution of technology rapidly increasing, there are ways to help OCR read handwriting — but it is still not 100% accurate. Image quality also plays a huge factor in the accuracy of OCR. The higher the quality of your picture, the higher the quality of OCR accuracy.

Pro: Indexing Time

When talking about OCR within a document management system, indexing is usually the term associated with it. Indexing is the process of classifying information that describes a document and allows you to easily search and find said document once it has been digitized. When you take all your paper documents and scan them, in order for them to be searchable you have to put them through some sort of indexing process.

There are two primary types of indexing used in document management systems today - indexing for content searching and indexing for database searching. For content searching documents, they must be processed through an OCR step before the text will exist to be indexed. The OCR technology allows for you to search the text by words found within the document. For example if you are indexing a case study about [“document scanning software”](#) you would be able

to search it by typing in keywords within the document. However, since OCR technology is not 100% accurate, make sure the document management system you are using has a review process in place to ensure that all text is correctly recognized.

Since the OCR does a lot of the work up-front, it potentially saves your company time in the beginning which could be beneficial depending on how fast you need to find these documents.

Con: Search Time

Now that we have talked about the benefits of OCR when it comes to indexing time, search time is a little more manual. It will take your company longer to search for the document than it did to index it. This is because you are searching for a document based on the content within it. Since many documents will have similar words and key phrases, you would have to go through many documents just to find the one you need.

This is not ideal for documents that need to be retrieved timely and regularly. However, if your company has a lot of documents that are not used or retrieved regularly it could be beneficial to deploy content searching as a 'last ditch' document retrieval method.

Pro: Database Records

Although it has limited uses there are instances where OCR is beneficial. Probably the best use of OCR within a document management system is to help you create the indexing data that will go into your structured database record. For example if you are recording a client's first name, last name, account number, etc. to go

into the database, there are certain ways that OCR can help to make this task less manual.

One way is through form scanning. With Optix, form scanning can select certain areas of a form - often called "zone OCR." The result of the OCR step is used to create a new database record. The new record and the scanned file are then stored in Optix or input into a new workflow. Using this technique, piles of paper forms can be quickly and digitally made available to your organization. This way the entire document doesn't need to be indexed, just the selected fields. By only indexing selected areas the database search allows you to get very precise results within seconds. With content search, however, you may come back with hundreds or thousands of results based on your search. That means your company would manually have to search long lists of documents to find the correct one. Form scanning is mainly useful for companies that handle multiple invoices, forms, etc. often.

What is Next?

Now that you have a better understanding of the pros and cons of OCR within a document management system, you may wonder how to take the next steps. With [Optix](#), we will help you start your process of becoming paperless with a document management system and also help determine what ways OCR can best benefit your company as a whole. Consulting with one of our professionals is the best way to determine what OCR needs your business has and the most efficient way to implement it. If you have any questions about OCR or how it works with a document management system, [give us a call today!](#)

