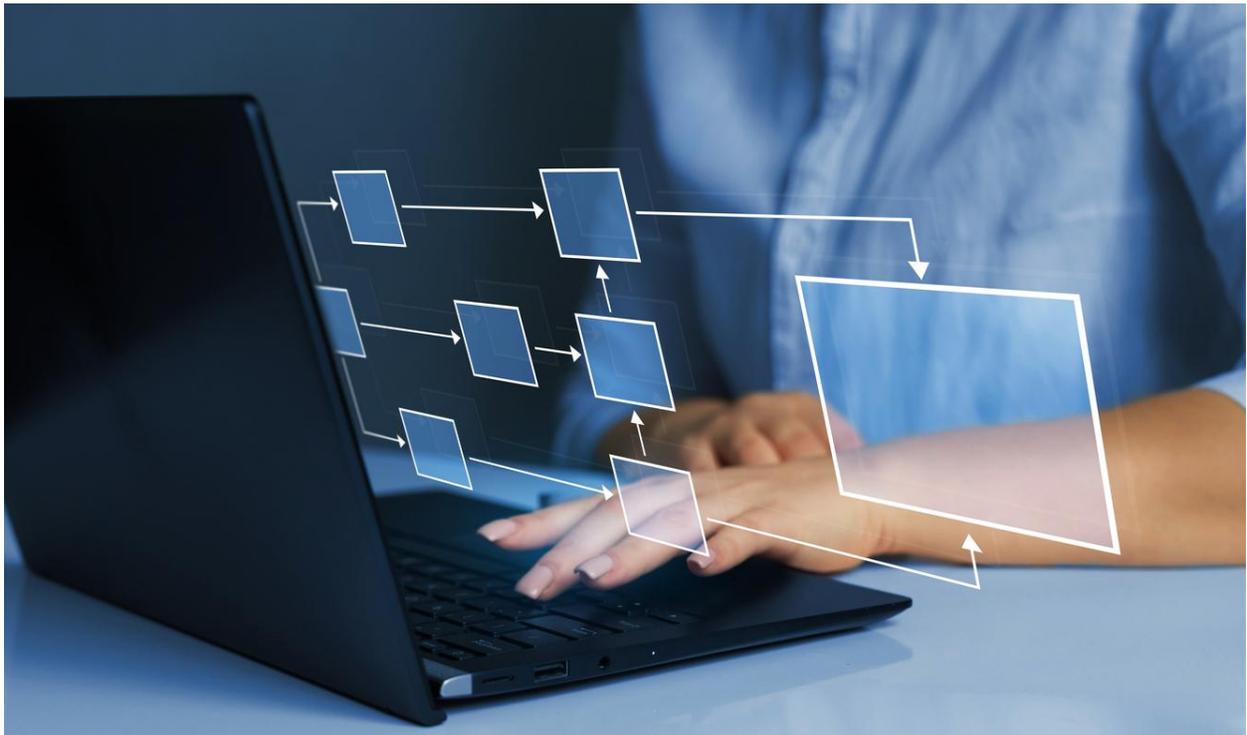




Using OCR to Streamline Forms Processing

May 16, 2022



As today's businesses continue to advance, it is important that technology keeps up with the business's growing demands and needs. OCR stands for [Optical Character Recognition](#) (OCR) and is a powerful tool to help businesses streamline processing forms.

Indexing documents can be a tedious and time-consuming task that requires a lot of manual work without the help of OCR.

Implementing OCR can greatly reduce workload and increase efficiency to serve your business's needs better. OCR in a document management system is usually used for [indexing features](#). At first sight, OCR seems like a simple and sleek way to quickly make your scanned documents searchable. These are some strong positive benefits, but overall there are some initial benefits and drawbacks to consider before implementing OCR within your document lifecycle. This blog focuses on exploring and discussing the strengths of OCR in streamlining forms processing to help you see if OCR is right for your business's needs.

General Methods

There are two general methods for indexing documents - structured and unstructured. Unstructured indexing uses OCR to attempt to capture the document contents in its entirety. With the text being fed into a text search database, OCR scans the document and analyzes it. Structured indexing identifies the most important/most searched data about a document and records this data in a traditional database.

Due to unstructured indexing relying on OCR, oftentimes when searching for a specific document, it can be more difficult than searching for a document in a structured indexing system. Unstructured indexing generates either too few or too many search results and can cause delays. Due to structured indexing summarizing a document, it creates an easier search experience for a user. There are many robust ways to [index your documents](#), making it tailored to your business needs.

Things to Consider

Attempting to OCR all of a document's content requires a large amount of computational time. More importantly, a simple search can miss documents that do not include the exact search terms used. Thus, the results are inaccurate and might not be what someone is looking for. In addition, users oftentimes have to wade through hundreds or thousands of text search hits to find the specific document they are looking for. Which increases the time someone has to spend looking for information and creates a greater work burden.

Searching a structured database is easier due to the program prompting a user to enter search term(s) that apply to all documents of a specific category (i.e., Legal, HR, etc.). A narrow search term is then required, such as a name or employee ID, which often finds the exact document needed in under a second in the database. No scouring hundreds or thousands of search results to find the correct one.

A more efficient way to deploy OCR is to selectively use it to extract the desired indexing data and then create a structured index record. This is called "Zone OCR" because an administrator identifies "zones" (rectangles) on a specific document type or form type (i.e., 1040 tax return, vendor invoice, etc.) and then pairs these with targeted database fields. This can greatly reduce or eliminate the need for manual document indexing.

To further explore the pros and cons of OCR, visit our "[The Pros and Cons of Optical Character Recognition](#)" blog to continue reading.

Optix OCR Solutions

Optix provides two major tools to help automate the indexing process using Zone OCR - Form Recognition and Zone OCR Templates to

efficiently process various form types. Form types are automatically recognized upon scanning. Then, the associated Zone OCR template that matches the specific form type is used to extract the required index data. If the data is missing, the portion of the scanned page corresponding to the defined zone is displayed along with any recognized text. This allows the scanning operator to quickly QC or correct the data before it is committed to the database.

Zone OCR can improve the accuracy over full-text OCR by allowing the administrator to specify the type of data being recognized on the form. For example, when recognizing numbers, Optix's built-in OCR engine has multiple recognition features such as only recognizing numbers, a limited set of letters, or a specific barcode type. By knowing in advance the type of data to be recognized, the administrator can improve the accuracy of the data obtained for each defined zone.

Optix is delivered with a built-in OCR engine and the ability to define and use Zone OCR templates during scanning. Optional modules add the ability to automatically recognize forms, automatically employ complex data extraction rules. For example, the OCR engine and Zone OCR templates can be instructed to find "Total:" in the desired zone and then extract the currency data to the right. Additionally, Optix's features can perform mark sense which determines whether or not a box is checked or a circle is filled in. With intuitive features, Optix creates a smooth system to accelerate your business's indexing process.

Get Started Today!

Understanding why OCR can transform your business is just the first step before implementing OCR into your indexing process. Clearly understanding which OCR will best serve your business's needs ultimately sets your business up for success. You can get started on

research to see which system suits your business's needs. Here at Mindwrap, we have industry-leading experts to help you get that conversation started and get your company moving forward in the right direction. [Contact us](#) today to get started.