square9
softworks

Your Guide to Mastering OCR
(Optical Character Recognition)

How to scale capturing information and simplify common business tasks

Table of Contents

square9 softworks

OCR's Role in a Digital World

Equinox's 2021 Global Interconnection Index predicted that by 2022, 65% of the world's GDP would be digitized. Are you part of that percentage? Are your competitors? The pandemic has undoubtedly helped fuel global shifts toward digital transformation. As a result, businesses needed solutions to help them work from anywhere at any time.

After adopting these new systems, businesses quickly realized a slew of additional benefits, such as increased productivity, reduced costs, and simplified, automated processes, to name a few. Many businesses also chose to permanently adopt remote and hybrid work environments, putting additional emphasis on their need for digital transformation solutions.

With so many previously paper-based business processes now being digitally transformed, accessing and leveraging the data stuck in those paper documents is of utmost importance.

And with OCR technology, you can!

Short for optical character recognition, OCR is a term broadly used to describe technologies that transform scanned images, readable only to the human eye, into data that computers and software solutions can use. Other technologies often associated with OCR include:

- **Intelligent Character Recognition (ICR)**
- **Optical Mark Recognition (OMR)**

Financial data can be lifted from invoices and moved to an enterprise resource planning (ERP) system, which can be used to make future business decisions.

square9
softworks

Using these technologies, businesses not only scan their documents to be digitally filed, but extract leverageable, manipulatable information from them. OCR solutions can even be used in tangent with artificial intelligence (AI) for even more accurate results. Using OCR and related technologies for the purpose of automatically scanning, indexing, and storing documents is often referred to as advanced document capture.

Sales data can be lifted from sales orders and used in a customer relationship management (CRM) system.

Thousands of organizations across industries are already utilizing capture automation to lift and index data so it can be stored in an enterprise content management (ECM) system. These systems help businesses store and retrieve documents such as invoices, contracts, employment records, sales orders, and more. They also help eliminate physical storage space, facilitate collaboration, automate tasks such as onboarding and invoice approval, and free up teams to work on bigger-picture projects. Document-heavy industries and departments often make great use of ECM and document capture solutions. Among these industries are:

- Education
- Healthcare
- Finance
- Manufacturing
- Distribution
- Accounts Payable
- Contract Management
- Accounts Receivable
- Taxes
- Human Resources
- And many more

OCR: A Wide-Range of Strategies for Varying Business Needs
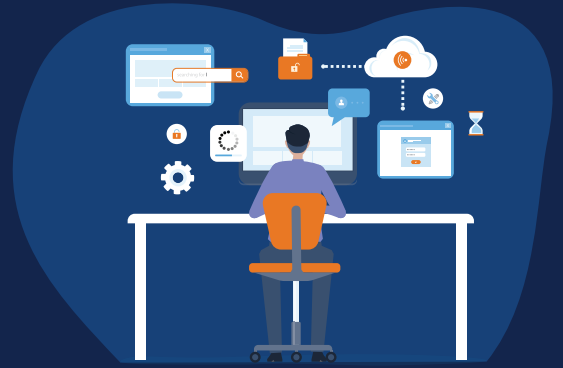
As a discipline, OCR implementation can be diverse. Its practice varies widely on a case-by-case basis and depends greatly on an organization's size, process maturity, technical sophistication, resources, and objectives for the solution's use. A common thread, though, is that the success of an implementation initiative depends to a great degree on an organization's understanding of its needs.

Capture Accuracy: Many factors play into the accuracy of an OCR solution, such as font size, type, and density; the use of irregular fonts and logos; and most notably, the source of the image. A PDF or quality image of the original file will almost always yield the best results but even in these cases, solutions are typically 99.7% accurate. Still, there are many cases where 100% accuracy is required for the task, such as:

> Sharing information into other line-of-business applications

> Indexing documents that need to be retrieved quickly

> Searching emails for data points to help assign tasks

When information needs to be 100% accurate, strategies that limit dependence on OCR or use formulaic checks to ensure accuracy are typically the most effective.

Example 1:
Suppose your company plans to index insurance forms and has a highly accurate source of related data. In that case, an OCR strategy may involve capturing one key index field to identify the document and pulling the other index fields from that data source.

Example 2:
On the other hand, if your company does not have an accurate reservoir of data to pull from, insurance information such as member ID and group number can be checked for accuracy by noting the format of these numbers and searching the document for other numbers with that format to see if they match.

square9
softworks

Age of Document: Older documents that are stained, torn, or faded may all impact capture accuracy.

Some capture solutions clean up scanned documents before beginning the capture process by adjusting for cockeyed scans; removing spots, lines, and other print imperfections; and compensating for stains, fading, and signs of age. If you work with high volumes of documents that contain these types of imperfections, it may be worth looking into systems that perform this cleaning process.

Fast and Frequent Retrieval: Accuracy becomes a high priority when documents need to be retrieved quickly or frequently.

Example 1:
If your business needs to reference invoices to respond to the needs and questions of your vendors, partners, or customers, having them readily available is your top priority. Accuracy is critical to locating this information quickly, as you would likely be using whatever index field is available on-hand.

Example 2:
Inversely, documents that are being captured because of mandated retention rates and potential audits are usually rarely referenced. Often companies have a day to provide these documents. In these less time-sensitive cases, the typical 99.7% accuracy will suffice since one incorrectly captured index field simply requires you to search with another field, after which you can quickly correct it for the future.

Capture Volume: The number of pages you'll need to capture regularly plays a vital role in determining your OCR strategy. While OCR speeds up the process of leveraging your information quite a bit, capturing large volumes of paper documents can still take time, especially if accuracy is a priority. In situations involving high capture volumes, several implementation strategies can be used.

Example:
An educational institution plans to capture student documents. In this case, all documents related to a specific student can be grouped and labeled with a bar code to ensure index fields common to each document are accurate.

There are other possible strategies, such as implementing validity checks and limiting reliance on OCR, but ultimately which strategy makes sense depends on your organization's needs and available resources.

square9
softworks

6 Key Questions for Determining a Strategy that Suits Your Business
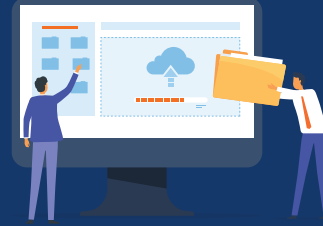
Additional factors can also affect your capture strategy, and it can be understandably difficult to keep track of these variables. So we've distilled these variables into key questions to ask when deciding on and implementing a capture solution:

1. Who or what department has control of the document format and layout?

Unusual fonts, special characters, words within logos and images, and other elements of a document's layout can all affect capture accuracy. Are the documents you plan to capture created internally or by vendors, customers, and other businesses? If you can control how your documents will be laid out, you may not need to implement as many measures to ensure accuracy.

2. What is your business's expected capture volume?

High volumes of documents need to be captured more quickly to keep up with demand. This affects the type of advanced capture you will need and could also affect accuracy. It is important to note that OCR is typically 99.7% accurate, and 0.3% of 1000 documents, for example, is much more noticeable than 0.3% of 100.

3. Is any of the data to be captured time sensitive?

If documents need to be captured and indexed quickly, it leaves less time to ensure the data is 100% accurate. This means that factors that affect capture accuracy need to be carefully considered, and an implementation strategy that limits the need for quality assurance may be ideal.

4. How accurately does this data need to be captured?

As discussed previously, not all captured index fields need to be 100% accurate. A solid quality assurance strategy for validating accuracy will change depending on the level of accuracy needed. Infrequently referenced documents, for instance, can be captured with far fewer validity checks as employees can simply use another index field to find them and correct the accuracy thereafter.
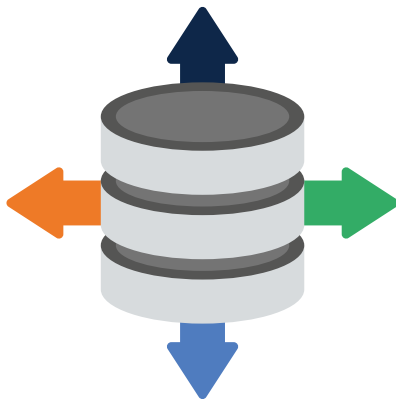
5. Does your business have a pure data source to leverage?

For documents that require 100% accuracy, one way to reduce reliance on OCR is to capture one key field and pull the other information from a pure data source. This can be any line of business application where you are sure data is accurate.

6. How often will documents need to be retrieved?

If documents won't be retrieved very often, then 100% capture accuracy is likely not required. In these cases, inaccurate indexing will still only happen about 0.3% of the time. In those events, another indexed field can be used to find the document, and the improperly indexed field can be promptly corrected.
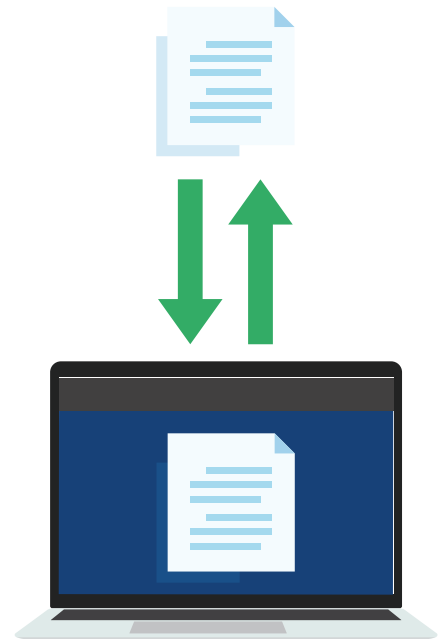
Good, Better, Best: Choosing the Capture Type that Works for You

Much like an artist selecting the right paintbrush for the job, so too must you select the document capture type that fits your unique needs. Different types of advanced document capture are designed for different workloads and circumstances. The correct capture type for your business depends heavily on the expected page volume and the layout consistency of the documents being captured.

Good:

KeyFree Indexing: With KeyFree indexing, employees can quickly index documents using a simple click and highlight method. This OCR tool speeds up indexing quite a bit but requires human input to populate each individual field. This type of capture also cannot send information to other business applications.

These distinct pros and cons make KeyFree Indexing a good option for indexing loads of 50 or fewer documents per day, but to meet other business challenges, a more comprehensive capture option may be needed.

Better:

Templated Capture: Templated document capture can be a highly effective solution when capturing document loads of more than 50 pages, especially when those documents use a set layout.

Since layouts are less likely to change with internal documents where your company controls the formatting or with documents from long-time vendors, templated capture is beneficial for internal processes or processes with set business partners.

Templated capture allows users to map set areas on a document to be captured and indexed, so in the future, information on documents with the same layout will be automatically lifted. If the system does not recognize the layout of a document, the system will request an employee to map it for future use. And unlike KeyFree Indexing, Templated Capture can send data to other business applications, helping your business make more informed decisions.

Best:

Unstructured Capture: Unstructured document capture can lift large loads of 50 or more documents a day without needing to map layouts.

This makes unstructured capture highly effective in situations involving high vendor turnover or where document formats constantly change. Unstructured Capture can also send data to other business applications for use in other processes.

Better Than Best:

Unstructured With AI: Industry-specific solutions may even use artificial intelligence in unstructured capture to learn from previous errors and improve accuracy. The most common solutions of this type are used for vendor invoices, receipts, and accounts payable automation.

square9
softworks

An Assortment of Implementation Methods: Finding the One that Meets Your Unique Needs

Simple OCR

SIMPLE OCR

The goal of a simple OCR strategy is to achieve 100% accuracy by limiting a process' reliance on OCR. The strategy typically captures a single data point that can be used to identify the entire document quickly. Once the system knows what document is being captured and what data is needed to index it, that information can be looked up in other lines of the business applications containing completely accurate data.

Using this system limits potential inaccuracies to a single data point and ensures that fewer documents need to be checked by human eyes. To illustrate, if that one data point is inaccurately captured, the capture system won't be able to find any information associated with it and will notify an employee that an error has occurred. This ensures that the only documents that need to be checked are those containing errors. The main challenge to implementing a simple OCR strategy is that a pure data source guaranteed to be accurate must be available to use and contain the information needed.

Complex OCR

In contrast to simple OCR, Complex OCR strategies capture each individual index field and use a series of checks to ensure accuracy. Examples of the potential checks include matching captured totals on an invoice against the numbers being summed and matching a captured name against other document areas where that name would occur. Using this method, it is possible to catch most of the inaccuracies that can occur with capture, but some index fields may not have a way to implement these checks, making human quality assurance important for 100% accuracy.

COMPLEX OCR
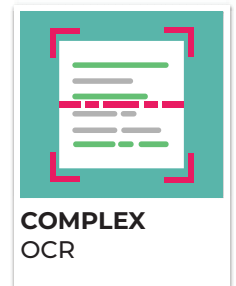
Situational OCR

SITUATIONAL OCR

There are situations where neither simple nor complex OCR makes sense, and a more niche approach is required. Examples include when documents need to be archived in bulk within a limited time. With large document loads affecting quality assurance time and the total number of estimated errors, a system that assists accuracy and saves time is essential. To achieve this, barcode cover sheets can be placed on each packet of documents containing a set of index fields. This way, fields that would be repeated on multiple documents, such as names and identification numbers, are guaranteed to be accurate, and each document is guaranteed to be searchable with those fields.

square9 softworks

How OCR Enhances the Big Picture and Helps Your Business

Increase Accuracy: Manually keying in information from paper and PDF is far more likely to lead to inaccurate data; after all, even the best employees are still human. When trying to transcribe high volumes of documents at a reasonable speed, people will undoubtedly make mistakes, and hours of work can understandably fatigue a person as well. OCR empowers employees to focus on catching the few inaccuracies that slip through the system.

Save Time: Whether indexing documents for archival or keying information into a business application, manually transcribing information from paper or PDF is tedious and time-consuming. OCR speeds up this process tremendously, allowing employees to put their time towards more business-critical projects.

Eliminate Double Data Entry: Using OCR systems to index documents and fill data into other business applications saves employees from having to key this information twice. This reduces paid employee hours spent keying in data and also reduces human involvement in the process, improving accuracy.

Reduce Costs: Using OCR with an ECM solution can save businesses from recurring costs such as document storage space, paper, ink, and toner. When combined with the reduced paid employee hours spent transcribing data and performing document-related tasks, businesses can see a significant financial return on investment.

Scale your Business: OCR frees up several essential business resources, including employee time, office space, and finances. With these resources now readily available, you can put them towards efforts to grow your business.

Improve Business Decisions: Make well-considered decisions using OCR to ensure that business-critical documents are correctly indexed and keep the information updated in other business applications. As a result, you can ensure employees and managers have the information they need to move forward on a decision that needs to be made.

Accelerate the sharing of information: With previously paper-based information now actionable in digital form, employees can securely search for and share this information by accessing well-indexed documents or by pulling data from other up-to-date lines of business applications.

square9
softworks

OCR Success Stories

Kuna Food Services Scales Up Their Business

Kuna Food services is a privately owned food service company distributing wholesale to restaurants, institutional facilities, and retail locations. Amidst a period of rapid growth, Kuna Food's needed a solution to manage the company's invoices, purchase orders, and bills of lading.

Using GlobalCapture to automate the flow of invoices and checks into their system, Kuna saves 20 hours a week. GlobalCapture also spotted accounting errors almost instantly, bringing users directly to the error for attention.

> Before, we were manually scanning and filing hundreds of invoices, but now that we have GlobalCapture, we can scan the same amount of paper, and OCR goes to work for us.
>
> **Tim Boyle** IT Manager
> Kuna Food Service

IntraVAT Switches Platforms for Optimal Accuracy

IntraVAT is a small subsidiary of the LEMAN group that files value added tax returns for businesses based out of European countries. Although IntraVAT began its digital transformation with a different capture solution, that solution was only 50% accurate. This, combined with a cyber attack, led the company to search for a better option.

IntraVAT's Square 9 solution was able to aid the company in ways that previously required 3 separate solutions. GlobalCapture provided IntraVAT with a 40% accuracy increase for their particular documents taking their capture accuracy from 50% to 90%. IntraVAT was also very happy with the modular pricing options and GlobalCapture's uncapped page limit.

> "Usually you pay per page, which makes a very big difference for us because we scan 3000 pages per day.
>
> **Linda-Sophia Danielsen**, CEO of IntraVAT

square9 softworks

![Square 9 Softworks logo]

Find out why businesses are turning to Square 9 for their OCR needs.

At Square 9, we offer an award-winning platform of solutions for assisting businesses of all sizes to boost efficiency, save time, and reduce costs. With a customer renewal rate of over 96%, Square 9 is the company that businesses have come to rely on. Our customers and their success are important to us, so we offer free e-learning to all active customers ensuring they get the most out of their solutions.

Visit Square 9 to get a no-obligation discovery session and see how OCR solutions can help improve your business processes and save your business time.

Email: info@square-9.com
Phone: 203-789-0889
www.square-9.com/products/document-capture-automation