Optical Character Recognition (OCR) for Invoices

IN THIS ARTICLE

Optical character recognition, or OCR, as we know it was first developed in 1974 by Ray Kurzweil.

His company developed the first omni-font OCR, OCR that could work with

virtually any font, as a way to assist the visually challenged by scanning information and converting it to digital audio.

However, it didn't become popular until the early 1990s when it began to be used to digitize newspapers and other historical documents.

Though early versions of most OCR tools had accuracy issues, the technology has been perfected over the years, with many OCR software applications now offering near-perfect accuracy when reading and extracting data from documents.

Before OCR technology existed, the only way to create a digital format of a document was to use manual data entry, which often resulted in inaccurate data and numerous typos.

Today, OCR is used in businesses for a variety of reasons.

One area that has benefitted from using OCR technology is accounts payable, where OCR invoice processing can be used to extract data from invoices directly into an accounting software application without the need to enter invoice data manually.

What Is OCR?

Optical character recognition, or OCR, is text recognition and data extraction technology.

It works by first reading data from printed text or handwritten documents and then translating that information into a usable format that can be edited.

In simpler terms, OCR uses a scanning device to copy or read text from a document, extracting the necessary data which then is translated into a format that can be used for a variety of purposes.

For years, OCR software was commonly used to extract data from historic documents and turn them into editable PDF documents.

But in recent years, businesses have increasingly turned to OCR technology to help automate common business processes in accounts payable, including reducing the amount of data entry needed for invoice processing.

What Is OCR Technology Commonly Used For?

OCR is one of many tools that can provide an automated invoice processing solution while also helping to automate your accounts payable department.

And with more businesses moving towards automation, there are several reasons why OCR technology can be used:

- Reducing or eliminating data entry by automatically scanning invoices, purchase orders, and receipts
- Scanning PDFs or other documents into a format that is easily accessible and editable
- Archiving historic data such as magazines and newspapers into a searchable format
- Depositing checks electronically
- Mail sorting for delivery
- Translating text from one language to another

What Is OCR in Accounts Payable?

Commonly used for automating the invoice process, OCR technology can read and extract invoice data such as invoice numbers, product codes, prices, and due dates into a usable format, eliminating the need to enter the information manually.

While using OCR alone can save time and increase accuracy, there are some instances where data cannot be read or may be extracted and saved erroneously.

However, when paired with artificial intelligence tools such as machine learning, OCR technology becomes more valuable, helping to increase accuracy and streamline workflow.

What Is Invoice OCR?

Invoice OCR is used to extract data from vendor and supplier invoices. OCR works by identifying characters on the invoice that need to be recognized. The process uses two distinct algorithms:

- ## Pattern Recognition

  OCR uses a database that contains a wide variety of fonts and font types, allowing it to recognize data on typed and handwritten documents. Pattern recognition also works to analyze various elements that may be included in the document, such as images and tables.

- ## Feature Extraction

  To further extract data, OCR also uses feature extraction, which drills down on essential pattern recognition to further identify specific characters.

  Feature extraction is used to identify features specific to a letter, such as the ability to recognize the difference between an F and an E for more accurate data capture.

How Does Invoice OCR Work?

Invoice OCR is used to automatically extract valuable data from invoices that are received from vendors and suppliers.

There are two ways that OCR can be used, either by manually scanning in invoices or with an invoice automation solution, which will automatically scan and extract data directly into your accounting system.

For general purposes, these are the steps that OCR typically uses to process invoices, regardless of whether they're received electronically, via email, or by regular mail.

1. ## Invoice Is Received

   The invoice is received from the vendor or supplier. If you're still receiving vendor invoices via email, fax, or regular mail, you'll need to manually scan the invoice into the system.

   If you're using an automated AP system, manual invoice scanning is not necessary since the invoice will be automatically scanned once received.

2. ## OCR Reads the Invoice

   OCR reads the invoice, extracting valuable information such as invoice amount, vendor and supplier name, and payment due date into a readable format. OCR technology used a two-color version of the document scanned.

   Before the document can be read, it must first be reduced to a black-and-white image. The image is also cleaned up during this step, removing any image spots or lines.

3. ## OCR Checks Each Element Using Pattern Recognition

   OCR checks each image, character, or word one at a time, using pattern recognition to identify each character, comparing it against a database of

fonts, characters, and letters.

After pattern recognition, it moves to feature extraction, which finds the best match for the previously scanned characters.

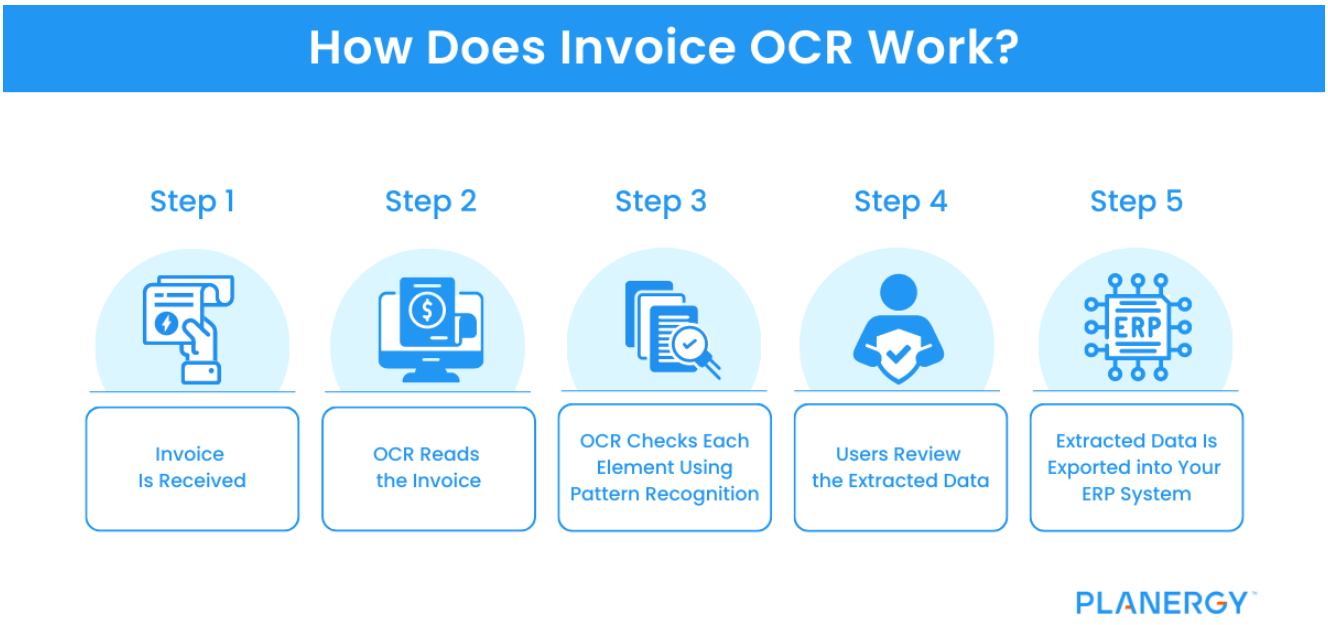
4. Users Review the Extracted Data

Users review the extracted data. Whether you're using OCR technology alone or it's part of a bigger automated AP system, you'll need to perform an initial review of the extracted data to ensure that it's accurate.

When combined with machine learning, the OCR engine will 'remember' any corrections or additions you make, so that future scans will be more accurate.

5. Extracted Data Is exported To Your ERP System

Extracted data is exported into your ERP system or accounting software application saving significant time on manual data entry.



How Accurate Is OCR Scanning?

Though OCR technology has improved in recent years, accuracy in OCR scanning is still a concern. there are currently no OCR resources available that can accurately recognize text with 100% accuracy.

Though many companies offering optical character recognition tout accuracy rates of up to 99%, the 1% is something that needs to be considered before opting to use an OCR solution for invoice data extraction and invoice processing.

1%, or even 0.1%, does not sound like a significant problem for accuracy but if the errors are on important characters in the scan this can be a bigger problem.

Inaccuracy on relevant fields, like price and quantity, impact the field-level confidence score.

The high accuracy rates are also normally based on best quality documents, high quality scans or digital PDFs. If you are working with lower quality documents the accuracy rates will come down significantly.

How To Improve OCR Accuracy?

The best way to boost OCR technology is to combine the use of OCR with other automated solutions that use artificial intelligence or machine learning. But there are ways to increase accuracy with OCR technology.

For invoice scanning many of these will be out of your control but you could always ask your suppliers to help where possible.

- ## Use High Quality Documents

  Document type makes a difference, so be sure to pay attention to the documents you're scanning.

  Documents that have a higher resolution and clear contrast such as black text on a white background will result in more accurate data extraction.

- ## Use Consistent Document Sizes

  Try not to vary the types of documents that are scanned.

  For example, OCR technology that is commonly used to read and extract data from word processing documents may have a lower accuracy rate if you scan in odd sized documents such as receipts.

  Of course, the technology should be able to handle both, but you may need to spend a little extra time reviewing documents to ensure accuracy.

- ## Use Consistent Layouts and Fonts

  Try to use standardized invoice formats and fonts. While this may not be possible for all documents, it can help maintain accuracy levels.

- ## Improve the Quality of Documents That Will Be Scanned

  For more accurate invoice data capture, pay attention to the quality of the documents being scanned.

  If text is too light, or there are various marks on the document, the result may be less than perfect.

Why Should You Digitize Invoices?

Digitizing invoices is an important part of accounts payable digital transformation that allows you to streamline the entire AP process and automate invoice processing, from invoice receipt to automated payment.

There are advantages to digitizing invoices for both sellers and buyers.

Sellers can reduce the invoice processing time dramatically by creating digital invoices instead of printing, stuffing envelopes, and mailing them to customers.

Customers can also save time by receiving digital invoices from their vendors and suppliers, which eliminates the need to enter invoice data manually, while

improving accuracy.

Digital invoices can also expedite invoice routing and approval time, while eliminating lost, misplaced, or destroyed documents.

Taking the time to work with your vendors and suppliers to institute technology that works for both of you is the ideal solution.

For example, if you have an automated AP system, but your suppliers are still sending you paper invoices, you'll be limited in the amount of technology you can utilize.

Another benefit of digitizing documents is the ability to store them electronically, eliminating the need to file documents. A digital filing system also makes it easier to later retrieve these same documents.

And even though digitizing invoices will not completely eliminate the need to review them for accuracy, the time needed is minuscule compared to the time spent processing invoices and other related documents manually.

How Often Should You Scan Invoices?

Smaller businesses with less volume should find scanning invoices once per week sufficient, while larger businesses may need to scan invoices at least twice per week to stay current.

If you use OCR technology along with other AI resources such as robotic process automation (RPA) or machine learning, scanning invoices will be part of the automation process, eliminating the need to scan invoices as they arrive.

What Are the Benefits of OCR?

There are several benefits to using OCR in the accounts payable process, particularly when used in conjunction with other AP automation tools. These benefits include:

Improved Accuracy

Using OCR technology, while not perfect, can provide more accuracy than manual data entry.

Even if OCR technology is used as a stand-alone solution, it can increase accuracy, particularly if a business typically processes a large number of invoices.

- ## Saves Time

Ardent Partners State of ePayables 2022 states that the time required to process a single invoice was 10.9 days.

OCR technology allows users to quickly scan an invoice, which is then read and exported to an ERP or accounting software application, thus eliminating time-consuming data entry, reducing the processing time to 3.71 days.

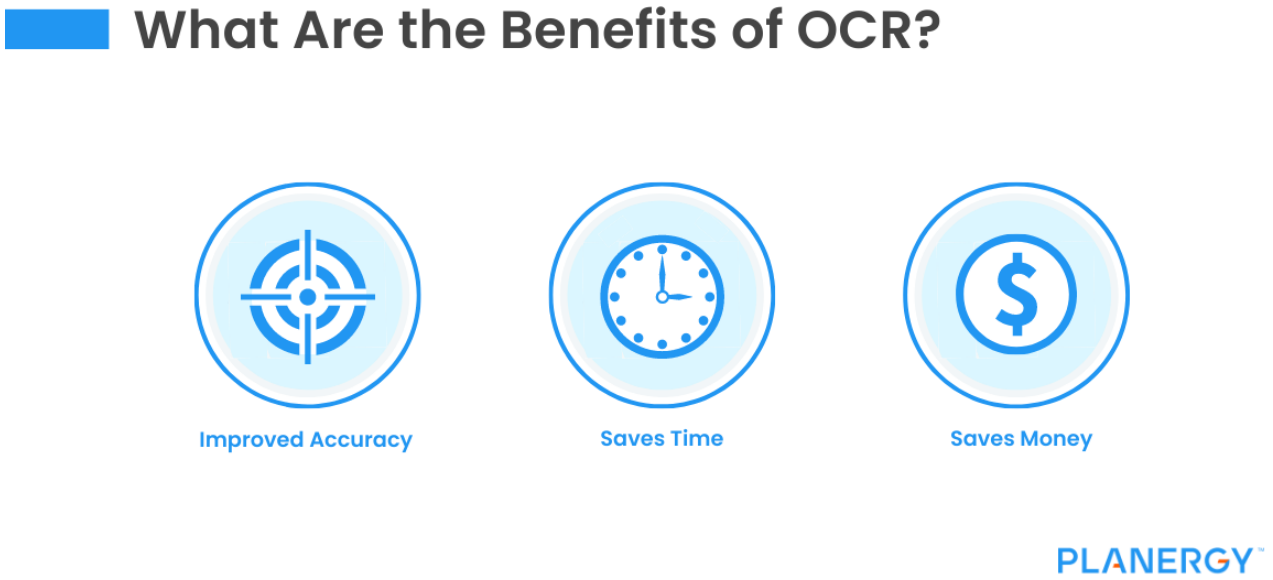
- ## Saves Money

In 2022, the average cost to process a single invoice was $10.18 when using manual AP processing.

Every hour spent processing an invoice costs money, whether it’s the AP clerk’s salary when entering invoices manually, the time spent copying invoices, or the time spent filing them.

Using OCR allows you to speed up the process considerably, allowing your employees to spend their time on more productive tasks, and with complete AP automation, invoice processing cost is reduced to $3.23 to process a single invoice.

-



What Are the Limitations of OCR?

For all of its benefits, there are limitations to OCR, particularly if it’s not used in conjunction with other AP automation tools.

- ## Still Requires Manual Checks for Accuracy

  Accuracy is both a benefit and one of the limitations of using OCR. While OCR technology offers better accuracy than entering invoices manually, it still needs to go through a validation process, where you can detect and correct errors.

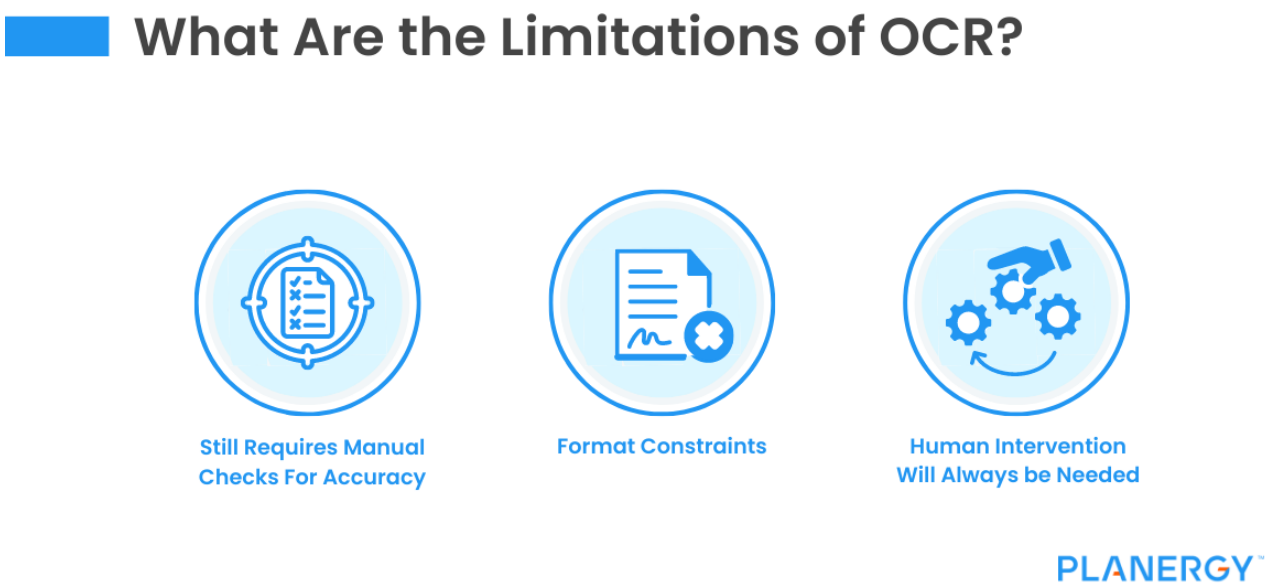
- ## Format Constraints

  While OCR technology performs well when reading and extracting data from standard documents, it may not be able to extract data from small or handwritten text, blurred or unclear text, illegible fonts, individual line items, or data from tables.

- ## Human Intervention Will Always Be Needed

  If you’re looking for true AP automation, you’ll need to introduce other automation tools such as AI or machine learning, since OCR alone will

always need human review to ensure extracted data is accurate.



How Does AI Work With OCR in AP Automation?

Every automated AP system uses OCR technology. However, using OCR alone can only automate a small portion of the AP process.

For true automation, you'll need to look for an automated system that uses OCR in conjunction with artificial intelligence and machine learning, which is used to identify and remember invoice structure and patterns for the future.

Using only OCR, you can quickly scan in an invoice, check the extracted data for accuracy, and then forward the invoice to approvers or for payment. But when you add machine learning to the mix, the process becomes much more automated.

Machine learning does require some initial human intervention; learning how to read the information on an invoice each time one is processed.

For example, if you process an invoice for building suppliers, you'll initially supply the necessary details for accurate processing, including the payment amount, invoice number, vendor name, due date, and general ledger expense code.

Once that is completed, machine learning technology helps the application

remember where that invoice is expensed, and what fields need to be identified on each invoice going forward.

OCR and AI combined can also eliminate one of the more time-consuming AP tasks; 3-way matching.

How Can You Use AP Automation and OCR for 3-Way Matching?

Using a procure-to-pay software like PLANERGY that incorporates AP Automation with OCR and AI, incoming supplier invoices are received electronically, automatically scanned, and processed using both AI and OCR, which also digitally archived each document.

Later, invoices are matched against existing purchase orders and shipping receipts, with the process fully transparent for audit purposes.

This process eliminated the need to manually match documents, while automatically flagging any discrepancies.

The Best Way To Use OCR

OCR is an important part of a comprehensive AP Automation software that can save time and money by streamlining the entire AP process from invoice processing to payment.

When used alone, OCR can help reduce manual data entry and manage documents electronically, but for true automation, the best solution is to use OCR along with other AP automation tools such as artificial intelligence, robotic process automation (RPA), and machine learning.

What’s your goal today?

1. Use PLANERGY to manage purchasing and accounts payable

We’ve helped save billions of dollars for our clients through better spend management, process automation in purchasing and finance, and reducing financial risks. To discover how we can help grow your business:

- Read our case studies, client success stories, and testimonials.
- Visit our Accounts Payable Automation Software page to see how PLANERGY can automate your AP process reducing you the hours of manual processing, stoping erroneous payments, and driving value across your organization.
- Learn about us, and our long history of helping companies just like yours.

Book a Live Demo

2. Download our guide “Preparing Your AP Department For The Future”

Download a free copy of our guide to future proofing your accounts payable department. You’ll also be subscribed to our email newsletter and notified about new articles or if have something interesting to share.

download a free copy of our guide

3. Learn best practices for purchasing, finance, and more

Browse hundreds of articles, containing an amazing number of useful tools, techniques, and best practices. Many readers tell us they would have paid consultants for the advice in these articles.

Related Posts