Best OCR Software of 2021

by **Prithiv S** 3 months ago   9 MIN READ

Table of Contents

- Is there any free OCR software?

Why do you need an OCR Software?

Businesses are turning increasingly digital to ramp up growth, and OCR software has been a key solution in this context. Scanning & processing documents such as invoices, receipts, and images for valuable data has traditionally been a manual process fraught with errors and delays. OCR software solutions help businesses save time and resources that would otherwise be spent on data entry & manual verification. OCR software automate data capture from scanned documents/images and digitize the data in convenient and editable formats that fit into organizational workflows.

Modern OCR software are fast, accurate and can handle common document processing constraints such as poorly formatted scans, handwritten documents, low quality images/scans, and blemishes that would have traditionally required extended manual interventions. More and more organizations are automating document processing workflows to go paperless and leverage cloud-based digital solutions that improve bottom lines.

What is OCR & what does OCR software do?

OCR or Optical Character Recognition is a technology that identifies & recognizes text within scanned documents, photos or images. OCR software leverages this technology to extract data from PDFs or scanned documents by converting it into machine-

readable text data that can be edited & stored more conveniently for further processing. For a detailed explainer on OCR and its use cases refer to this guide.

OCR is also used in various other use cases such as extracting tables from PDFs, extracting text from images or extracting text from PDFs or other non-editable formats.

Today, OCR software is used for automated data entry, pattern recognition, text-to-speech services, indexing documents for search engines, cognitive computing, text mining, key data and machine translation among various other applications. These tools can convert any scanned documents, PDFs or image types into xml, xlsx or csv files.

The best OCR Software for your business

Let's look at some of the best OCR software available on the market.

Nanonets

Nanonets is an AI-based OCR software that automates data capture for intelligent document processing of invoices, receipts, ID cards and more. Nanonets uses advanced OCR, machine

learning image processing, and Deep Learning to extract relevant information from unstructured data. It is fast, accurate, easy to use, allows users to build custom OCR models from scratch and has some neat Zapier integrations. Digitize documents, extract data-fields, and integrate with your everyday apps via APIs in a simple, intuitive interface.

Nanonets Intro

How does Nanonets stand apart as an OCR software?

Pros:

- Modern UI
- Handles large volumes of documents
- Reasonably priced
- Ease of use
- Cognitive capture of data - resulting in minimal intervention
- Requires no in-house team of developers
- Algorithm/models can be trained/retrained
- Great documentation & support
- Lots of customization options

- Wide choice of integration options

- Works with non-English or multiple languages

- Almost no post-processing required

- Seamless 2-way integration with multiple accounting software

- Great API for developers

Cons:

- Can't handle **very high** volume spikes

- Table capture UI can be better

Nanonets online OCR & OCR API have many interesting use cases that could optimize your business performance, save costs and boost growth. Find out how Nanonets' use cases can apply to your product.

Get Started

ABBYY Flexicapture

FlexiCapture is a stable, scalable document imaging and data extraction software that automatically transforms documents of any structure, language or content into usable and accessible business-ready data.

ABBYY FlexiCapture for Invoices - Demo Video

Pros:

- Recognizes images very well
- Easy to store hard copy result in system
- Integrates well with ERP systems
- Automates data extraction from documents (to an extent)

Cons:

- Initial setup can be difficult and complex
- Automatic processing of invoices not set up
- No ready-made templates
- Difficult to customize
- No resources available
- Could have better integration with RPA solutions
- Low accuracy with low resolution images/documents
- *Batch verifications are held up even if there's an error just in a particular section*
- *Line item error messages pop up even for items that should be skipped*
- *RESTful API is not available in the on-prem version*

ABBYY Finereader

ABBYY FineReader PDF is an OCR software with support for PDF file editing. The program allows the conversion of image documents into editable electronic formats.

Processing Documents with ABBYY FineReader Server – Demo Video

Pros:

- Keyboard-friendly OCR editor for manual corrections
- Exceptionally clear interface
- Exports to multiple formats
- Unique document-compare feature

Cons:

- Lacks full-text indexing for fast searches
- Requires a learning curve
- *Pricing can be prohibitive*
- *Inability to view the history of document changes*
- *Can't merge several files into one*

- *Might require some post-processing*
- *The UI could be overwhelming at first*
- *Slow to process big files*

Need an OCR software for image to text extraction or PDF data? Looking to convert PDF to table, or PDF to text? Check out Nanonets online OCR in action!

Get Started

Kofax Omnipage

Omnipage is a powerful OCR software that can handle automation for high-volume corporate OCR tasks. This tool specialises in table extraction, line item matching, and smart extraction.

Pros:

- Has a robust set of tools for enhancing images
- Highly accurate

Cons:

- UI not intuitive
- Configuration for AP Automation is not straightforward
- API integration can be improved

IBM Datacap

Datacap streamlines the capture, recognition and classification of business documents to extract important information from them. Datacap has a strong OCR engine, multiple functions as well as customisable rules. It works across multiple channels, including scanners, mobile devices, multifunction peripherals and fax.

Pros:

- Configures complex applications in data capture
- Scanning mechanism
- Ease of use

Cons:

- Very little online support
- UI could be more intuitive
- Setup can be cumbersome
- Slow
- Creating a customized flow isn't straightforward
- Batch commits take time

Start using **Nanonets for Automation**. Try out the various OCR models or request a demo today. ***Find out how Nanonets' use cases can apply to your product.***

<div align="center">

Try Now

</div>

Google Document AI

One of the solutions in the Google Cloud AI suite, the Document AI (DocAI) is a document processing console that uses machine learning to automatically classify, extract, enrich data and unlock insights within documents.

Pros:

- Easy to set up
- Integrates very well with other Google services
- Storage of information
- Speed

Cons:

- AI modules lack proper documentation
- Customization of existing modules and libraries is hard
- Not suited for Python or other coding languages
- Outdated API documentation
- Expensive
- Not suited for hybrid cloud deployments
- Not suited for use cases that require custom AI algorithms

AWS Textract

AWS Textract automatically extracts text and other data from scanned documents using machine learning and OCR. It is also used to identify, understand, and extract data from forms and tables. For more information check out this detailed breakdown of AWS Textract.

Pros:

- Pay-per-use billing model
- Ease of use

Cons:

- Can't be trained
- Varying accuracy
- Not meant for handwritten documents

Want to scrape data from PDF documents, convert PDF table to Excel or automate table extraction? Check out Nanonets PDF scraper or PDF parser to scrape PDF data or parse PDFs at scale!

Get Started

Docparser

Docparser is a cloud-based document processing and OCR software that can automate low-value tasks and workflows for businesses.

Pros:

- Easy setup
- Zapier integration

Cons:

- The webhooks occasionally fail
- Requires some deal of training to pick up the parsing rules
- Not enough templates
- Zonal OCR approach - can't handle unknown templates
- UI could be better
- Slow to load pages
- *Documentation could be better*

Adobe Acrobat DC

Adobe provides a comprehensive PDF editor with an in-built OCR functionality.

Pros:

- Stability/compatibility.
- Ease of use

Cons:

- Expensive
- Not an exclusive OCR software
- *Heavy on the system*
- *Takes up a lot of space on the hard disk*
- *Difficult to integrate with services like Sharepoint or Dropbox*
- *Requires an Adobe Creative Cloud license.*

Klippa

Klippa provides automated document management, processing, classification and data extraction solutions to digitize paper

documents in your organization.

Pros:

- Fast setup

- Great support

- Great API for developers

- Clear and concise API documentation

- Links well with accounting programs

- Competitively priced

- Integrations

Cons:

- OCR recognition can be better

- Limited template customizations

- Limited white-label customizations

- Bulk adjustments not supported

- The VAT is often not displayed correctly

- The app crashes often

- Can't train the OCR model

- The selection process isn't straightforward as there are a lot of options

Nanonets online OCR & OCR API have many interesting use cases that could optimize your business performance, save costs and boost growth. Find out how Nanonets' use cases can apply to your product.

Get Started

Other notable mentions include **Veryfi**, **Readiris**, **Infrrd**, **Rossum** & **Hypatos**.

Here's a quick comparison of all the OCR software listed above across some crucial OCR software features & parameters:

	Fields Captured	Key-value Pair Extraction	On-premises Deployment	Table Capture	Intuitive UI	Ease of Customization	3-Way Matching	Transparent Pricing	Free Trial	Integrations	On-chat Support	TOTAL
Nanonets	1	1	1	1	1	1	0	1	1	1	1	10
Rossum	1	1	0	1	1	0	1	0	1	1	1	8
Klippa	1	0	0	1	1	0	1	0	1	1	1	7
Docparser	1	0	1	0	0	0	1	1	0	1	1	7
Kofax Omnicapture	1	1	1	1	0	0	1	0	0	1	0	6
Veryfi	1	1	0	1	0	0	0	1	1	0	1	6
Google Document AI	1	1	0	1	0	0	0	1	1	0	0	5
AWS Textract	0	1	0	1	0	0	0	1	1	1	0	5
ABBYY Flexicapture	1	0	1	0	0	0	1	0	0	1	0	4
Adobe Acrobat DC	0	0	1	0	0	0	0	1	1	1	0	4
ABBYY Finereader	0	0	1	0	0	0	0	1	1	1	0	4
Readiris	0	0	1	0	0	0	0	1	1	1	0	4
IBM Datacap	0	0	1	0	0	0	1	0	0	0	0	2

How does Nanonets stand apart as an OCR software?

Nanonets OCR software is **easy and flexible to set up**, requiring just about 1 day. The automation handles **unstructured data** without much difficulty and the AI also handles **common data constraints** with ease. Information from **documents with imperfections & blemishes** is extracted quite easily. It handles **multi-page invoices** and identifies **multi-line items** with ease; something that most legacy and modern OCR tools fail at. Nanonets **customizes column headers** allowing it to process complex invoices more efficiently. Nanonets' AI also ensures

a **high accuracy** while processing documents requiring minimal rework or revision.

The benefits of using Nanonets go just beyond better accuracy, experience and scalability. Here are 8 reasons that highlight the unique Nanonets advantage:

1. **Training & working with custom data** - Most OCR software out there are quite rigid on the type of data they can work with. Nanonets isn't bound by such limitations. Nanonets uses your own data to train models that are best suited to meet the particular needs of your business.

2. **Easy to use & flexible** - Adapting Nanonets for your specific business needs is easy and straightforward. From creating custom OCR models & retraining them to adding new fields & handling integrations, Nanonets can handle it all.

3. **Learns & retrains continuously** - Businesses often face dynamically changing requirements and needs. To overcome potential roadblocks, Nanonets OCR software allows you to easily re-train your models with new data. This allows your OCR model to adapt to unforeseen changes.

4. **Customize, customize, customize** - Nanonets can capture as many fields of text/data that you like and present it in any desired fashion. Captured data can be presented in tables or line items or any other format of your choice with custom validation rules. Always remember that Nanonets is not bound by the template of your document!

5. **Requires almost no post-processing** - While most OCR software simply grab and dump data, Nanonets extracts only the relevant data and automatically sorts them into intelligently structured fields making it easier to view and understand. This does away with a lot of time spent in revision and verification.

6. **Handles common data constraints with ease** - Nanonets leverages deep learning & object detection techniques to overcome common data constraints that greatly affect text recognition and extraction among other OCR software. Nanonets AI can recognize and handle handwritten text, images with low resolution, images with new or cursive fonts and varying sizes, images with shadowy text, tilted text,

random unstructured text, image noise, blurred images and more. Traditional OCR software are just not equipped to perform under such constraints; they require data at a very high level of fidelity which isn't the norm in real life scenarios.

7. **Works with non-English or multiple languages** - Since Nanonets focuses on training with custom data, it is uniquely placed to build a single model that could extract text from documents in any language or multiple languages at the same time.

8. **Requires no in-house team of developers** - No need to worry about hiring developers and acquiring talent to personalize Nanonets API for your business requirements. Nanonets was built for hassle-free integration. You can readily integrate Nanonets with most CRM, ERP or RPA software.

Is there any free OCR software?

Apart from the professional cutting-edge OCR solutions mentioned above, there are free OCR software that do the job to an extent. Running on open-source OCR engines (like Tesseract), these free solutions help convert photos, PDFs, TIFFs or scanned documents into editable digital text formats. While they might not be able to process elaborate business documents at scale, they are adequate for extracting text from simple documents with straightforward formatting.

These free OCR solutions either come as web-based applications, standalone software that need to be installed on various platforms, or as a side feature in a full-fledged document editing service. Please note that free OCR software regularly fail to process handwritten documents, multi-column tables, long line items, or low quality images/scans.

Here are some free OCR options for your consideration:

- OnlineOCR.net
- FreeOCR.
- SimpleOCR
- GOCR
- Office Lens
- English OCR
- Easy Screen OCR
- A9t9
- Photo Scan
- Capture2Text
- Adobe Scan
- OCR Using Microsoft OneNote
- OCR With Google Docs

Update May 2021: this post was originally published in January 2021 and has since been updated.

Here's a slide summarizing the findings in this article. Here's an alternate version of this post.