

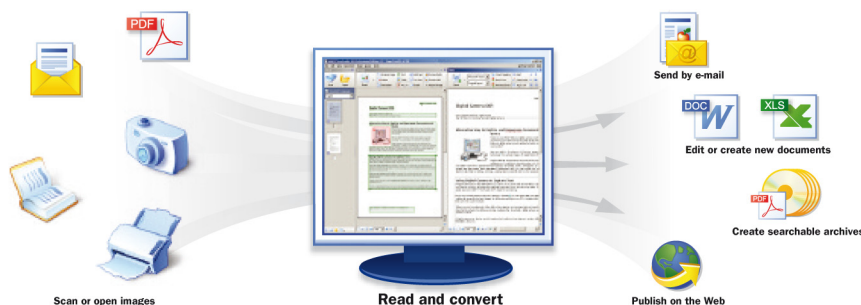
# What is OCR?



## What is Optical Character Recognition - OCR?

OCR is a technology that enables you to convert different types of documents, such as scanned paper documents, PDF files or digital camera images into truly **editable and searchable data**.

Imagine you've got a paper or a PDF document – such as magazine article, brochure or contract – and you need to edit or use the content. You can either **retype the document or use OCR software**.

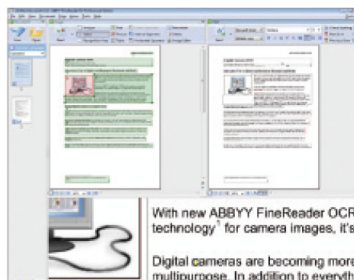


OCR software transforms a picture of letters and words from a scanned or photographed document into real letters and words and enables you to access and edit the content of the document. Such a professional OCR application is **ABBYY FineReader**.

## How does ABBYY FineReader OCR work?

The OCR process consists of three basic steps:

- **Scan document or open image (or PDF)**
- **Recognition** = change picture into editable text
- **Save as DOC, RTF, XLS, PDF, HTML, TXT etc. or export data directly to Microsoft® Word, Excel or Adobe® Acrobat.**



The program starts the image recognition process with an analysis of the document structure. It divides the page into elements such as blocks of text, tables and images. Then it detects lines, individual words and letters. Each 'character picture' is compared to a set of patterns and the program defines the highest probability of a certain letter. Supported by sophisticated algorithms and the new **'Adaptive Document Recognition Technology' (ADRT™)** FineReader determines the position of each letter in a word, each word in a line as well as the properties of all pages and their relation to each other.

In addition, FineReader provides dictionary support for 36 languages. The program compares detected words with the words in one or several of the dictionaries, which ensures an **even higher recognition accuracy**. The recognition results of ABBYY Fine-Reader OCR exceed 99% on standard documents.\* The layout of each page with all formatting elements such as tables or columns will be exactly reproduced. The logical structure of the whole document will be reconstructed, headers and footers, footnotes and page numbers will appear in native Microsoft Word formatting elements.

\* OCR recognition results depend on the quality of the analysed document. The recognition is typically more difficult on low resolution or damaged document

## Customer Benefits

Advanced, powerful OCR software such as ABBYY FineReader allows you to **save a lot of time and effort** when electronically processing or archiving documents.

Extract quotes from books or magazines and use them to create your case scenarios and presentations without retyping. Capture text from web banners, posters, timetables or books in libraries with a digital camera and then use the captured information for your own purposes. Or use OCR software to create searchable PDF archives from your paper documents.

The entire process of data conversion from original paper documents, images or PDFs takes minimum time and effort and the final recognised document looks just like the original! **You will never need to retype a document again!**