

## ABBYY® FineReader Engine 12

Integrate AI-powered OCR features into your applications.

# Why ABBYY FineReader Engine?

The AI-powered software development kit ABBYY FineReader Engine allows you to integrate ABBYY's multilingual text recognition, PDF conversion, data extraction, barcode reading and document classification technologies into your applications. This way, you can quickly create highly accurate text extraction and document conversion solutions - with low programming effort.

## New functions for your application

ABBYY technologies enable software applications to transform printed text and tables into MS Word or Excel, convert large document archives into PDF/A format, and extract individual data from scans, photographs or screens.

## Text recognition and data extraction

The toolkit enables integration of ABBYY's technologies for recognition of machine-printed and hand-printed texts, detection of checkmarks, barcode recognition, as well as technologies for extraction of data from business cards and from the machine readable zones (MRZ) in ID documents.

## Powerful PDF processing tools

The SDK allows document conversion into many types of PDF and PDF/A formats. When importing PDFs, the technology can preserve PDF properties and inject text into areas without text layer. XML data can be extracted from incoming PDF/A-3 documents — and added to the PDF/A-3 files during the PDF export.

## AI-based document classification

The AI-based document classification utilizes OCR, Natural Language Processing and a pre-trained Convolutional Neural Network. This allows to quickly train the classification module on own documents and implement precise and granular document classification.

### VERSATILE OCR SDK FOR YOU

- Add document processing functionality to your software and save on programming.
- Create applications turning scans, photos and screenshots into formats accessible by computer systems.
- Convert paper documents into editable text, searchable PDF, PDF/A, XML, HTML, RTF, ODT, e-book, Microsoft Office, vCard and many other formats.
- Enhance solutions for banking, insurance, legal, government or healthcare sectors with document processing and information extraction functionalities.
- Enable machine vision or RPA systems to read data from screens or industrial displays.
- Offer software for many platforms — Windows, Linux, Mac, Cloud and embedded systems.

# Advantages of ABBYY FineReader Engine

Adding OCR, ICR, barcode reading, business card data extraction, recognition of MRZ areas, and other functionalities via an SDK integration is quicker than own development. This way, you can market your products faster and your clients can benefit from your solutions earlier.



Within its comprehensive set of technologies, ABBYY FineReader Engine provides the highest number of OCR languages in the market. The high recognition results are achieved through deployment of intelligent algorithms that leverage artificial intelligence technologies, advanced image pre-processing tools, and integrated dictionaries. Untypical words, new languages or fonts can be added through pattern training.



Development with ABBYY FineReader Engine is fast and efficient. With extensive documentation, provided code samples, user interface components, ready-to-use document processing profiles with optimized parameters, and technical support, even demanding OCR projects can be quickly implemented.



The ABBYY SDK offers outstanding scalability, availability and flexibility. Processing pages in parallel on multiple CPU cores can significantly increase the speed. The flexible licensing allows integrating the technology into own cloud-based OCR services or using it in virtual and other highly scalable environments - for fast conversion of extensive document archives into PDF/A or targeted data extraction from individual fields.



With ABBYY FineReader Engine, the recognition process is in your hands. Developers have full access to internal recognition results – information about character coordinates, advanced attributes such as font and formatting details, and access to hypotheses about word and character recognition can be used in external systems.

## Broad range of input formats

The SDK can process PDFs, images such as photos, screenshots or scans, as well as documents in different Office formats. They can be acquired directly from memory, uploaded from storage or scanned via the TWAIN or WIA interface.

## Variety of saving formats

The SDK offers many export options, such as TXT, RTF, DOCX, XLS(X), CSV, HTML, HTML5, ODT and PPTX that can be directly edited, E-book formats EPUB and FB2, XML, the library standard ALTO XML, XPS, vCard for business card data, as well as many types of PDF and PDF/A formats.

## Exact layout reconstruction

Leveraging the ABBYY Adaptive Document Recognition Technology and Machine Learning algorithms, the SDK allows you to create searchable and editable documents that exactly match scanned or photographed originals.

## Data from business cards and machine readable zones (MRZ)

ABBYY BCR technology allows exporting business card data directly into contact management systems. ABBYY tools for extracting information from MRZ in ID documents can be used in customer onboarding systems for fast entry of personal information.

# ABBYY FineReader Engine: Recognition process step by step

The document recognition process consists of several steps, starting with document input and ending with the export of the results. Sophisticated algorithms are applied during each step to deliver highest recognition accuracy and precise layout reconstruction.

## Step 1 Input

Files in Office formats, images or screenshots can be uploaded or acquired from the memory stream. All types of PDFs can be processed - their annotations, metadata, bookmarks and other data can be kept.

## Step 2 Image Pre-processing

Pre-processing tools optimize the images prior to the text recognition step. To achieve high-quality recognition results, images are rotated, cropped, de-skewed and binarized. Distortions are corrected and backgrounds are filtered out.

## Step 3 Layout Analysis

Document analysis is performed to detect text areas and collect information about the document, its structure and the layout of each page. Choose from several document analysis modes or manually define text recognition blocks.

## Step 4 Recognition

Machine-printed text in over 200 languages, hand-printed text in more than 120 languages, many different fonts, writing styles and language combinations can be recognized. Barcodes and checkmark values can be extracted.

## Step 5 Verification

Internal recognition results, such as character coordinates, fonts and formatting are accessible and can be used to implement automated correction or manual verification (for Windows, ActiveX user interface components are available).

## Step 6 Export

The recognition results are delivered according to requested settings. These can range from individual field values in TXT or CSV format to completely reconstructed PDF, PDF/A or Word documents — with their internal links and original formatting.

## NEW Support for processing in the Cloud and on Virtual Machines

With an enhanced licensing system, your document processing and data extraction applications can be deployed on premise, in the Cloud on Amazon AWS and Microsoft Azure or in virtual environments e.g. VMware Workstation and ESXi, Docker Containers and Oracle VM VirtualBox.



## NEW Input of Office documents, recognition of MRZ areas & new OCR languages

Ability to input documents in Office formats, recognition of Machine Readable Zones in ID documents together with new recognition languages such as Farsi, Georgian, basic math formulas, technical preview of Burmese and improved quality of Japanese and Chinese OCR further expand ABBYY's leadership.



# Specifications and Licensing

With flexible licensing policies, ABBYY FineReader Engine offers the best combination of tools and pricing options for a variety of projects.



## Individual and Flexible Licensing

The licensing can support different scenarios, e.g. processing certain number of document pages (pages per month, pages per year, one-time projects), local deployment, usage within a computer network, in the cloud, and others.



## Modular Configuration

Ad-on license modules allow enhancing the standard functionality with individual options such as additional recognition languages, MRZ recognition, classification or specific document processing options.



## Software Maintenance and Upgrade Assurance

With the Software Maintenance and Upgrade Assurance, you receive regular updates and bug fixes, continuous technical support and new versions of FineReader Engine free of charge.



## Professional Services

ABBYY offers Professional Services that can support your development and integration projects. If you wish to speed up your project by leveraging ABBYY development expertise, please contact us.



## Development & System Requirements

- ABBYY FineReader Engine for Windows supports the COM standard and can be used in C/C++, Visual Basic, .NET, Delphi, Java and tools supporting COM components. It can be adapted for use in scripting languages like VBS, JS, and Perl.
- ABBYY FineReader Engine for Linux provides C/C++ native API and a Java wrapper.
- ABBYY FineReader Engine for Mac provides C/C++ native API.

For technical details about supported operating systems, cloud and virtual environment support, as well as details about recognition languages and barcode types, please visit [www.abbyy.com/ocr-sdk/technical-specifications](http://www.abbyy.com/ocr-sdk/technical-specifications).

**Note:** The availability of features in the Windows, Linux and Mac versions might differ. Please [contact ABBYY for details](#).

## TECHNOLOGIES FOR INTEGRATION

### OCR

Multilingual recognition of machine-printed text in 210 languages with Latin, Cyrillic, Greek and Gothic alphabets; Arabic, Chinese, Farsi, Hebrew, Japanese, Korean, and Vietnamese.

Support for standard printer, dot-matrix printer, typewriter, OCR-A, OCR-B, CMC7, MICR and receipt type of text.

### ICR

Multilingual recognition of hand-printed letters and digits in 126 languages and 22 handwriting styles.

### OBR

Recognition of 1D and 2D barcodes printed in any angle — including damaged barcodes.

### OMR

Recognition of many types of checkmarks — single or grouped checkmarks, checkmarks in cells or on a white background, and checkmarks that were corrected.

### Business Card Recognition

Ready-to-use extraction of data from business cards in 27 languages.

### MRZ Recognition

Ready-made tools for extraction of personal information from ID documents.



To learn more about the software development kit ABBYY FineReader Engine please, visit [www.abbyy.com/ocr-sdk](http://www.abbyy.com/ocr-sdk). If you have additional questions, contact your local ABBYY representative listed under [www.abbyy.com/contacts](http://www.abbyy.com/contacts) or use the online contact form [www.abbyy.com/ocr-sdk/#request-demo](http://www.abbyy.com/ocr-sdk/#request-demo).

[WWW.ABBYY.COM](http://WWW.ABBYY.COM)