

## ABBYY FineReader Engine 12 | Product Information

### Why ABBYY FineReader Engine?

The software development kit ABBYY FineReader Engine allows you to integrate ABBYY's multilingual text recognition, PDF conversion, data extraction and barcode reading technologies into your applications. This way, you can create highly accurate, scalable and flexible document conversion and text extraction solutions – while saving on time and own 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 text, detection of checkmarks, barcode recognition, and business card recognition.

### 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 PDF/A-3 documents - and added to the PDF/A-3 files during the saving step.

### Document classification

The enhanced module for document classification utilizes Machine Learning and Natural Language Processing together with OCR technologies, and allows to implement unattended sorting of documents into pre-defined categories.

### 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.
- Deploy document processing solutions for banking, insurance, legal, government and healthcare sectors.
- Enable machine vision or RPA software to process 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 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 by intelligent algorithms using broad set of character classifiers, 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 is known for its outstanding scalability, high availability and flexibility. Processing pages in parallel on multiple CPU cores significantly increases the speed, while the backup network license manager provides for continuous processing even in case of hardware failure. The flexible licensing supports cloud-based conversion of extensive document archives into PDF/A as well as extraction of individual characters.



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.

### Input: Broad range of input formats

The SDK can process photos, scans and images in many formats and from many different sources. Images and PDFs can be acquired directly from memory, uploaded from storage or digital cameras and scanned via the TWAIN or WIA interface.

### 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.

### Export: Variety of saving formats

The SDK offers many export options, such as TXT, RTF, DOCX, XLS(X), CSV, HTML, 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.

### Ready-made business card recognition

ABBYY technology for recognition of business cards automatically splits several business cards on one scanned page, extracts data from individual fields and exports them into the vCard format to manage contact 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 the highest quality recognition accuracy and layout reconstruction.

## Step 1: Input

Documents can be directly scanned or photographed. Images or screenshots can be uploaded or captured 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 for the text recognition step. For best possible 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 | Farsi, Burmese, improved Japanese OCR

Expanding ABBYY's leadership in the number of recognition languages, the latest version of ABBYY FineReader Engine offers improvements in the recognition of the Japanese language, adds Farsi as a new OCR language and Burmese OCR in technical preview mode.



## Technologies for Integration

### OCR

Multilingual recognition of machine-printed text in 208 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.

## Specification 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 license can support different scenarios, e.g. to process certain number of document pages or characters, utilize defined number of CPU cores, deploy applications locally, within a computer network, or in the cloud, and others.



### Modular Configuration

Ad-on license modules allow enhancing the standard functionality with individual options such as additional recognition languages, 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 for your SDK version, 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 the Windows, Linux and Mac versions and the respective features might differ. Please contact ABBYY directly for details on availability.



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).

This software includes ABBYY® FineReader® Engine 12 recognition technologies. © 2017, ABBYY Production LLC. ABBYY, FINEREADER and ABBYY FineReader are either registered trademarks or trademarks of ABBYY Software Ltd.

All product names, trademarks and registered trademarks are property of their respective owners. Windows® is a registered trademark of Microsoft Corporation in the United States and other countries. The registered trademark Linux® is used pursuant to a sublicense from LMI, the exclusive licensee of Linus Torvalds, owner of the mark on a world-wide basis. Mac® and OS X® are trademarks of Apple Inc., registered in the U.S. and other countries. Datalogics®, The DL Logo®, PDF2IMG™ and DLE™ are trademarks of Datalogics, Inc. Adobe®, The Adobe Logo®, Adobe® PDF Library™, Powered by Adobe PDF Library logo, Reader® are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States and/or other countries. Portions of this computer program are copyright © 1996-2007 LizardTech, Inc. The software is based in part on the work of the Independent JPEG Group. Portions of this software are copyright ©2011 University of New South Wales. Unicode support: © 1991-2013 Unicode, Inc. Intel® Performance Primitives: Copyright © 2002-2008 Intel Corporation. Portions of this software are copyright © 1996-2002, 2006 The FreeType Project [www.freetype.org](http://www.freetype.org). WIBU, CodeMeter, SmartShelter, and SmartBind are registered trademarks of Wibu-Systems. All rights reserved.

All other trademarks are the property of their respective owners.