PRODUCT BROCHURE

ABBYY

ABBYY FineReader Engine

The most comprehensive OCR SDK for software developers

The AI-powered software development kit ABBYY FineReader Engine allows you to integrate ABBYY's multilingual text recognition, PDF conversion, 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.

Whether you are a software vendor, system integrator, or an enterprise company developing your own IT systems, ABBYY OCR SDK will help you create highly accurate text and data processing applications.

Create desktop or server applications for Windows, Linux, or Mac and deploy them in the Cloud or on Virtual Machines. The diverse OCR features can add value to applications within many areas, such as DMS, ERP, RPA, insurance, banking, healthcare, legal, and machine vision.

Comprehensive set of recognition technologies

With the help of convolutional neural networks, end-to-end recognition, and deep learning for language models, the SDK offers highly accurate recognition of machine-printed text (OCR) and hand-printed text (ICR) that can be performed as full-page or field-level recognition. Also, it provides recognition of barcodes (OBR) and checkmarks (OMR).

Powerful PDF processing tools

The SDK offers document conversion into searchable PDF and PDF/A formats. When importing PDFs, the missing text layer can be injected while PDF properties are preserved. XML data can be extracted from imported PDF/A-3 files as well as inserted when saving to PDF/A-3 formats.

Artificial intelligence (AI) and machine learning (ML)

Powerful AI- and ML-based technologies combined with ABBYY's ADRT and other technologies detect the logical structure of a document and identify its formatting elements, such as table of contents, headers, footers, fonts, and font styles to precisely re-create the original document.

Multi-core CPUs and parallel processing

When converting many pages such as complete document archives or books, developers can leverage the SDK's flexible and scalable architecture. By using multi-core CPUs and processing images in parallel on multiple threads, the OCR steps can be performed significantly faster.

Advanced image pre-processing functions

Image pre-processing increases the recognition accuracy by optimizing the image for OCR. Even low-quality images can deliver best OCR results after de-skewing, rotation, distortion correction, text line straightening, page splitting, adaptive binarization, ISO noise reduction, and other image correction steps.

Recognition profiles for fast implementation

Predefined processing profiles contain pre-set parameters for the best OCR results within frequent recognition tasks. They can be used in place of manually fine-tuning the recognition system and provide for faster SDK integration.

Document classification with machine learning

Advanced classification algorithms in FineReader Engine leverage machine learning and OCR technologies, and enable applications to automatically categorize documents into different classes.

Support for cloud and virtual environments

In addition to on-premise deployment, software developers have the option to deploy their applications in virtual environments or host their offerings on cloud platforms like Microsoft Azure or Amazon Web Services.

User interface components

Developers can use ABBYY's ActiveX-based visual components to easily integrate user interface into variety of applications such as ERP, ECM, or DMS systems, providing their users with the possibility to validate documents, access the recognition results, and edit text and directly in the application.

Sample code library

A comprehensive set of code samples provides instructions on how to use the OCR SDK, illustrates different FineReader Engine technologies, and can be used for development projects. Each library section contains code snippets accompanied by systematic instructions on how to perform the relevant task.

Detailed SDK documentation and support

Detailed documentation helps to streamline integration. Developers have access to additional information through ABBYY's Help Center. In addition to these resources, ABBYY technical support is available to you during your free trial.

Compare documents function

The new functionality "Compare Documents" allows you to compare two versions of the same documents and detect differences in their content. The documents can be in different formats, e.g., Microsoft Word, PDF, TIF, JPEG, and any other of the supported input documents formats.

Feature overview

	Windows	Linux	Mac
Recognition			
BBMP, DCX, GIF, JBIG2, JPEG, JPEG 2000, PCX,scanned PDF, digitally created PDF incl. password protection, PDF/A, PNG, TIFF, (For Windows versiononly: WIC-compatible, WDP, DjVu)	+	+	+
Text documents (doc, docx, rtf, htm / html, txt, odt); tables (xls, xlsx, csv, ods); presentations (ppt, pptx, odp)	New	Planned	–
OCR - Recognition of machine-printed text	210 languages	208 languages	208 languages
Arabic, Chinese, Japanese, Korean, Hebrew, Thai, Vietnamese, Gothic OCR	+	+	+
Burmese (technical preview) and Farsi OCR	New	New	New
Georgian OCR	New	Planned	–
OCR for simple mathematical formulas	New	Planned	–
Japanese OCR	Improved	Improved	Improved
Chinese & Japanese OCR	Improved with ML	–	–
Recognition of dates, times, and addresses in Japanese and Arabic	Improved	Improved	Improved
Recognition of Machine Readable Zones in IDs (MRZ)	New	Planned	–
Ready-to-use data extraction from specific document types	Business cards	Business cards	Business cards
ICR - Recognition of hand-printed text	126 languages	Planned	–
OMR - Checkmark recognition	+	Planned	–
OBR - Barcode recognition	+	+	+
Image Pre-Processing Tools			
Auto-cropping and auto-splitting of dual pages	+	+	+
Filtering of color stamps and marks, noise removal, and local contrast improvement	+	+	+
Adaptive binarization	+	+	+
Camera OCR	+	+	+
Easy Development			
Scanning utility	Improved	+	+
Native 64-bit support	+	+	+
Code samples and predefined processing profiles	+	+	+
Bulit-in recognition on multi-core CPUs	+	+	+
Multithreading capability	+	+	+
Parallel processing of many single-page documents	+	+	+
Parallel processing of multi-page documents	+	+	+
Ready-to-use components for creating user interface	+	+	+
Layout Reconstruction			
Document layout reconstruction, incl. internal structure and formatting	+	+	+
Detection and recreation of balanced columns of text	New	New	New
Detection of tables and layout reconstruction	Improved	Improved	Improved

	Windows	Linux	Mac
Export Settings			
Broad range of export settings, such as RTF, DOCX, TXT, ODT, CSV, XLS(X), HTML, PDF, PDF/A, XML, PPTX, ALTO, EPUB, FB2, vCard, image formats	+	+	+
Export to HTML5	New	New	New
Faster export to XML with non-deskewed coordinates	Improved	Improved	Improved
Direct export of list elements to XML	New	New	New
ALTO 3.1 export format	New	New	New
Direct export to memory	+	+	+
Export to XPS	+	–	–
PDF Export			
Export to PDF 2.0 and PDF/UA	New	New	New
Export to PDF/A-1a, PDF/A-1b	+	+	+
Export to PDF/A-2 (a,b,u), PDF/A-3 (a,b,u)	New: PDF/A-2b, PDF/A-3b	New: PDF/A-2b, PDF/A-3b	New: PDF/A-2b, PDF/A-3b
Ability to use Unicode characters during PDF encryption	New	New	New
Export to Tagged PDF	Enhanced	+	+
Preservation of original annotations and bookmarks in PDF	+	+	–
Ability to detect text layer in PDFs and skip OCR	+	+	–
Fast export of PDFs with many pages	+	+	+
Technology Improvements			
Document Classification powered by ML and NLP	Enhanced	Enhanced	Enhanced
Access to information about individual recognition variants	New	New	New
Usage of artificial intelligence and machine learning for accuracy and layout reconstruction	New	New	New
Ability to open the PDF, DjVu, WIC, and WDP files from memory	+	PDF only	–
Ability to specify resolution for rasterisation during PDF opening	+	+	–
Full support for ZUGFeRD-compliant PDF (import & export)	+	+	+
Licensing			
Support for Cloud (Amazon EC and MS Azure) and virtual environments	New	New	-
Network licensing	+	+	-
Backup Network License Manager for higher fault-tolerance	+	+	-

+ already available in the version 11
– feature currently not available/ planned

ABBYY

To learn more about the software development kit ABBYY FineReader Engine please, visit www.abbyy.com/ocr-sdk
If you have additional questions, contact your local ABBYY representative listed under www.abbyy.com/contacts

abbyy.com