

## FineReader Engine 9.0

For Windows®



### The Most Comprehensive SDK for Recognition and Document Conversion

ABBYY FineReader Engine 9.0 is a comprehensive software development kit (SDK), which includes all the technologies needed for developing state-of-the-art intelligent document recognition and document conversion systems. In addition to OCR, ICR, OMR and Barcode recognition, it also offers image pre- and post-processing, field-level (zonal) recognition and PDF conversion.

As a result, corporate developers, VARs, and ISVs benefit from working in a single, flexible and expandable environment for easy integration with a variety of DMS/ECM applications.

#### Multiple technologies in one SDK

ABBYY FineReader Engine provides full spectrum of ABBYY technologies with functions that cover the following major business areas in a single platform: full-page text recognition, PDF conversion and field-level (zonal) recognition.

Moreover, ABBYY FineReader Engine intelligently identifies photographed documents and applies its adaptive recognition technology for camera images to ensure the best OCR results.

#### Expandable SDK and Technology Platform

ABBYY FineReader Engine delivers continuity in programming, maintenance and training. Developers can source a single SDK for different tasks. Additionally, plan future extensions of solutions and offer new services by working from the same platform.

#### Flexible for any type and size of application

Easily integrated with any type of software architecture ABBYY FineReader Engine can be used to build applications of any scale and complexity — whether or not it is a client workstation designed from scratch, a small

project for a few workstations, a server-based solution, or a large-scale project to be distributed among both servers and workstations.

#### Ease of deployment

ABBYY FineReader Engine offers easy access to core technologies and its COM API through development environments such as C/C++, Visual Basic, and Visual Studio.NET. Additionally, the Command Line Interface (CLI) allows a single command line call to set up a complete task.

#### Cost Effective

A modular architecture offers a variety of features as “add-on” modules, allowing developers to choose only the functions they need, while providing the option to add new functions at a later time.

#### ABBYY always in touch

Being an international company, ABBYY has local offices and partners' centers of competence all around the world. Thus ABBYY developers are sure to receive technical support:

- In time (within their time-zone);
- In a native language (from a local VAR);
- For any case (since ABBYY provides technologies of its own development).

### Product Benefits

#### Adaptive Document Recognition Technology (ADRT™)

allows recreating of document logical structure and formatting attributes including headers, footers, page numbers, fonts and styles and more.

#### Multi-core Recognition Architecture

facilitates the application speed of multi-core processors during image analysis and recognition.

#### Significant enhancements in Chinese, Japanese and Korean OCR

guarantee better accuracy and faster processing.

#### New MRC Compression

for PDF and PDF/A files significantly reduces the size of final documents.

## Key Functionality



### Image Manipulation and Pre-processing

ABBYY FineReader Engine 9.0 can receive images from three types of sources: scanning via TWAIN interface (including ADF support and manual input feeding), getting directly from memory, or opening from files. It supports major imaging formats, including multipage TIFFs, JPEG 2000, and works with black-and-white, grayscale and color images. It also can open PDF files, converting them into images using Adobe® PDF Library Technology. ABBYY FineReader Engine also may save original and modified images into various formats.

**ABBYY FineReader Engine can perform the following pre-processing functions to improve the recognition:**

- Automatically deskew images
- Split dual pages
- Despeckle image
- Texture filtering and adaptive binarization
- Auto-detection of page orientation (90, 180, 270 degrees)
- Manipulating with text color and background images
- Work with digital camera images
- Despeckling of an image in individual blocks (zones of an image)
- Data extraction from fields with various borders and frames

### Document Analysis

#### Field Level/Zonal Recognition

ABBYY FineReader Engine 9.0 delivers complete field level/zonal recognition capabilities to support key business processes such as forms processing, key-word classification, and keyword indexing. Powerful image pre-processing functions increase ABBYY FineReader Engine's ability to detect small zone areas of any quality, with any type of image nuance which may affect the recognition accuracy (i.e. underlined text, after-scanning garbage, spaces in the text, etc.). Key functionality for field level or zonal recognition includes multilingual OCR and ICR, OMR, and barcode recognition.

#### Document Analysis for Full Text Indexing

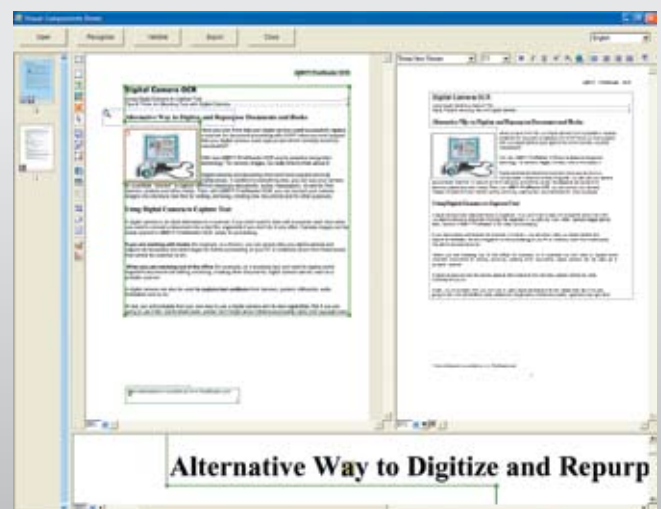
This function automatically detects and recognizes all text on documents including text embedded in pictures, charts, and diagrams. Developers may choose to use this mode of document analysis to extract exhaustive full-text information on documents needed for document index building (as in DMS, CMS, archiving systems).

### Visual Components

ABBYY FineReader Engine 9.0 includes Visual Components that represent a set of graphical user interface components. When creating own applications developers can give users direct but controlled access to recognition results and functions for validation or checking of documents. ABBYY FineReader Engine 9.0 Visual Components were created using ABBYY's expertise in end-user applications and are based on interfaces created for our award-winning ABBYY FineReader Professional Edition application.

**Visual Components in ABBYY FineReader Engine 9.0:**

- Scan Interface
- Document Viewer
- Image Viewer
- Text Editor
- Text Validator

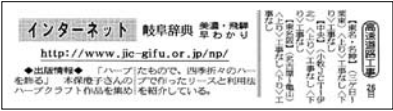




# Key Functionality

## Recognition OCR

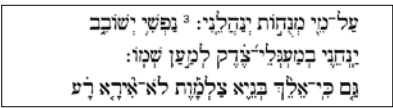
- Recognize up to a total of 195 machine print languages
  - 52 languages have dictionary/morphology support
- Recognition of multilingual documents
- Recognition of dot-matrix documents
- Recognition of typewritten documents
- Support of OCR-A, OCR-B, MICR (E13B) and CMC7
- Three OCR modes: fast, balanced and normal.
- ABBYY CameraOCR™
- **Chinese, Japanese, and Korean (CJK)** character recognition with added support for full-page export to PDF and RTF with vertical text and complex layout retention. Functionality enhancements in ABBYY FineReader Engine 9.0 include enhanced accuracy, faster processing and the ability to combine these hieroglyphic Asian languages with other languages.



- **Thai** words can be composed of four levels, with vowels written behind, over, under or around the consonants and diacritics located over and under the basic characters. In addition, there are no spaces between words. With the new technology, the SDK detects characters and separates text strings from each other to provide reliable recognition of Thai texts.



- **Hebrew** is written from right to left using the Hebrew alphabet, while numbers are written in the opposite direction, from left to right (most Hebrew texts today use European digits). In addition, texts in Hebrew often include words in English and other “left-to-right” languages.



## ICR

- Recognize up to a total of 113 handprint languages
  - 29 languages have dictionary/morphology support
- Recognition of hand-printed characters in various field borders and frames – underlined fields, boxes, comb-style fields, etc.



- Multilingual ICR
- 26 regional styles of hand-writing for supported ICR languages
- Supports Indian ICR digits that are used in Arab states.

## External Dictionary Support

In addition to internal ABBYY FineReader Engine dictionary types (standard, user, and regular-expression-based dictionaries), developers now have the ability to leverage external dictionaries, created based on their rules, inside the SDK.

## OMR

The OMR recognizes simple checkmarks, grouped checkmarks, model checkmarks and checkmarks with “corrections” made by hand. It delivers an accuracy rate of 99.995%.

## Barcode

- Recognition of various 1D barcodes, including EAN 13 supplemental barcode types and 17 1D barcode types. EAN 13 supplemental remains the standard in the publishing industry for encoding ISBN numbers on books.
- Recognition of 2D Barcodes: PDF417, Aztec, DataMatrix, QR Code
- Fast barcode extraction. This feature enables automatic finding and recognizing barcodes at any angle on a document. It works both for 1D and 2D barcodes.



## PDF Conversion PDF Input

- **Intelligent PDF processing.** FineReader Engine analyses internal information within the source PDF files such as annotations, metadata, text objects, font dictionaries and content streams and as a result enhances PDF performance and speed by efficient and accurate selection. When the text is embedded, examines the integrity of the text layer, and makes a decision as to whether or not to extract the text or apply OCR on a block by block basis.
- **Capture of internal PDF information,** extracting internal PDF links, hyperlinks and document properties such as: subject, author, title, and keywords.

## PDF Output

- **PDF security and encryption support** – the ABBYY FineReader Engine 9.0 platform supports a variety of PDF security settings, increasing its applicability for government agencies and other organizations demanding high security.
  - “Open File” password settings designed to prevent unauthorized access to a document.
  - Restriction of certain operations, such as printing, editing or extracting file content, by assigning permission passwords.
  - Support for the latest encryption standards.
- **Output in Tagged PDF format** – that can be “reflowed” to fit different page or screen widths. Ideal for use with handheld devices (PDAs) or screen readers typically used by visually impaired users.
- **Page size** – Ability to set the size for all pages of a output PDF file.
- **Links in PDF files** – Recreates hyperlinks within a PDF file.
- **PDF/A export** – export to PDF/A format which is recommended as a standard for long-term preservation of page-oriented documents (ISO 19005-1).
- **CJK to PDF export** – Enables conversion of documents in Chinese (both simplified and traditional), Japanese and Korean into PDF format.
- **MRC (Mixed Raster Content) Compression** – the size of the final document is up to 8-10 times smaller compared to JPEG compression, ideal when colour documents are scanned and processed.

# Features and Specification



## Special feature Development Platform Advanced Functionality

- **Voting API support** – gives developers additional information for using ABBYY FineReader Engine as one of the participating recognition engines in an external (or third-party) voting algorithm. Supplies recognition alternatives (or hypotheses) with relevant confidence levels for characters, words and inter-character separation.
- **“On-the-fly” tuning of core recognition** – ability to manipulate the Engine during the recognition process. Developers can influence the choice of hypotheses by inserting additional ranking criteria, which are used by the technology during the recognition to deliver the best result.
- **Code samples for common conversion tasks** – an additional resource for quick and easy tuning of the ABBYY FineReader technology. A set of code samples with ready-to-load profiles provides an optimal balance of speed and accuracy for particular tasks, such as conversion to searchable PDF, field-level recognition, archiving with imaging and indexing, and full-page conversion to RTF and HTML.

## Trial and Licensing

### Trial Version

ABBYY offers a fully functional trial version. To obtain an evaluation copy, please contact ABBYY reseller in your region.

### Developer Licenses

Grant the right to develop and test applications incorporating ABBYY FineReader technology.

### Runtime Licenses

Grant the right to distribute applications with ABBYY FineReader Engine 9.0 functions incorporated. Perpetual Runtime Licenses differ by functionality and the quantity of pages they allow to process per year. Server Runtime Licenses differ by number of CPUs that are used in the process.

**Add-on Modules for Runtime Licenses:** PDF (PDF/A) Export; XML Export; ICR & OMR; Asian Languages; 2D barcodes

## Pricing and Availability

The price depends on the number of add-on modules and the number of pages which can be processed. Please contact ABBYY certified value-added distributors and resellers in your region.

## Developer Environments

- Microsoft Visual Studio.NET (VB.NET, C#);
- Microsoft Visual Basic 5.0, 6.0;
- Microsoft Visual C++ 4.x and above;
- VB Script, and other scripting languages;
- Borland Delphi 2.0 and above;
- Any other development environment that supports COM and ActiveX objects correctly.

## System Requirements

- PC with Intel® Pentium®/Celeron®/Xeon™, AMD K6/Athlon™/Duron™/Sempron™ or compatible processor with a minimum clock speed of 200 MHz.
- Operating System: Microsoft® Windows® Vista, Windows Server 2008, Windows Server 2003, Windows XP, Windows 2000, and 64-bit versions of Windows Vista, Windows Server 2008, Windows Server 2003, Windows XP.
- Memory: 128 MB RAM plus an additional 110 MB for each recognition process.
- Hard disk space: 350 MB for full developer installation and 70 MB for program operation.

Please check [www.abby.com](http://www.abby.com) for other system requirements for network installation.

## OCR

- 52 languages with dictionary support
- 133 additional languages with Latin, Cyrillic or Greek characters
- 5 Asian languages with dictionary support: Chinese (Traditional, Simplified), Japanese, Korean, Thai
- Hebrew language with dictionary support
- 5 FineReader XIX languages with dictionary support
- 4 artificial languages
- 6 programming languages
- Simple chemical formulas
- Digits
- Tools for creating user-defined languages

## ICR

- 29 languages with morphology/dictionary support
- 84 languages without dictionary support
- Supports 26 styles of hand-printing of different countries and areas: European, American, Canadian, Russian, Japanese, Arabic and Thai

## Barcode Types

**1D:** Check Code 39, Check Interleaved 25, Code 128, Code 39, EAN 13, EAN 13 Supplemental, EAN 8, Interleaved 25, CODABAR (without checksum), UCC Code 128, Code 2 of 5 (Industrial, IATA, Matrix), Code 93, UPC-A, UPC-E and Postnet.

**2D:** PDF 417, Aztec\*, DataMatrix\*, QR Code\*

\* available – Q2 2009

## Image Input Formats

PDF, including PDF/Archive (PDF/A)  
BMP: black and white, gray, color  
PCX, DCX: black and white, gray, color  
JPEG, JPEG 2000: gray, color  
TIFF: black and white, gray, color, multi image  
GIF  
PNG  
DjVu

## Document Saving Formats

PDF, PDF/A, MRC (Mixed Raster Content) for both PDF and PDF/A  
DOC, DOCX  
XLS, XLSX  
PPT, PPTX  
RTF  
HTML  
TXT  
DBF  
XML  
CSV

### ABBYY Russia

Russia, Baltic States, Asia, Middle East, South America, Africa, Australia & NZ  
P.O. Box #49, Moscow, 129301, Russia  
Tel: +7 495 7833700  
Fax: +7 495 7832663  
sales@abby.com

[www.ABBYY.com](http://www.ABBYY.com)

**ABBYY®**