

## ABBYY® FineReader® Engine 8.1 Four-in One-SDK



### *Where to Use FineReader Engine*

FineReader Engine combines a full range of functions with the highest quality recognition, effective processing speed, and convenient development tools to help third party developers deliver the most powerful and efficient solutions for their customers. This makes it ideal for a variety of applications including:

- Archiving and document processing applications
- Control and verification systems
- Document conversion systems
- Fax processing applications
- Form processing applications (structured or unstructured)
- Content creation and management applications
- Digital mail-room applications
- Document sorting applications
- Web publishing systems
- Intranet archiving applications
- Media clipping solutions
- Reading or voice-playback systems

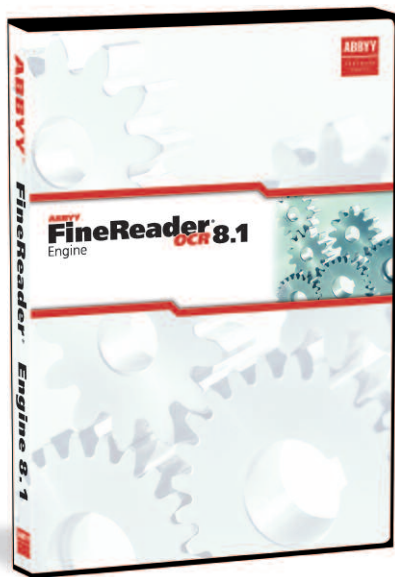
*ABBYY FineReader Engine is a Software Development Kit (SDK) for integrating ABBYY's core OCR (machine-print), ICR (handprint), OMR (check mark) and barcode recognition; PDF processing; fixed and semi-structured document processing technologies into external applications.*

### *Cost-Effective, Easy development*

#### **Benefits for Corporate Developers, VARs, and Third-Party Developers**

FineReader Engine makes it easy for developers to customise the technology to suit the needs of their particular application. Benefits include:

- **Single, expandable SDK and technology platform.** With the four document conversion and data capture operations in a single environment, 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 enough for any type and size of application.** Easily integrated with any type of software architecture 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 is to be distributed among both servers and workstations.
- **Ease of deployment.** 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 (CLEI)** allows a single command line call to set up a complete task.
- **Cost Effective.** A modular architecture and pricing model 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.
- **Easy licence management.** FineReader Engine's variable licencing scheme offers a single SDK and a single licence agreement for managing a variety of licencing options



## Four-in-One SDK

The most comprehensive SDK for document conversion and data capture, FineReader Engine provides full spectrum of ABBYY recognition technologies with functions that cover the following four major business areas in a single platform:

- full page text recognition,
- PDF conversion,
- field-level (zonal) recognition,
- and data capture.

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

### Full Page Recognition:

Delivers full text information for use in a variety of indexing and archiving functions. Highlights include:

- **Automatic Document Analysis (DA)** normal" DA mode with layout analysis for full page conversion tasks, plus two DA modes for the following tasks:
  - Full text indexing
  - Invoice pre-processing
- **Multilanguage, omnifont OCR and PDF conversion**
- **Multiple levels of layout retention**, including complete retention of document formatting such as columns, tables, text and pictures, paragraph styles, fonts, colour, etc.)
- **Export of detailed character information, export to XML** in 3 schemes (Native, MS Word XML, and ASCII XML)
- **3 pre-defined processing modes:** Fast, Balanced and Default/Accurate

### PDF Conversion

PDF opening and export functions for a variety of conversion and archiving functions.

#### ■ PDF Input

**Intelligent conversion scheme**, using internal information (annotations, meta-data, font dictionaries, and content streams) to analyse files and determine when to extract text layers and when to apply full recognition on a block by block basis. Ensures faster conversion and high accuracy.

**Capture of Internal PDF Information**, extracting internal PDF links and hyperlinks. Compliance with security setting of source PDF

**Adobe PDF Library technology**, included to ensure high accuracy and compatibility with current versions of the PDF format.

#### ■ PDF Output

**Security and encryption support**, supporting open and permission passwords, allowing restriction of print, editing, or extraction of content. Supports RC4-based 40-bit to 128-bit, and AES-based 128-bit encryption.

**PDF Export**, in multilingual, linearised, and "tagged" PDF formats. Saves "edited PDF" files. Four saving modes include "text under image" searchable PDF, image only, text over image, and text and pictures. Supports addition of PDF metadata, such as cross-links, hyperlinks, bookmarks, document properties.

**New in Version 8.1: PDF/ A Support**

## Field Level/Zonal Recognition

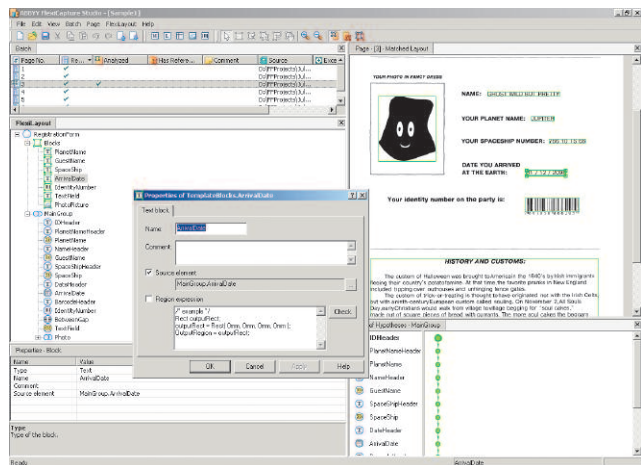
High quality recognition of zones, or snippets of images, containing valuable data Ideal for keyword classification and indexing.

- **Multilingual and digit ICR (hand print)**, recognising 22 handwriting styles (covering the pan- European and American writing styles), digits, letter and digit combinations.
- **Barcode recognition** with more than 16 popular 1D industrial types (New in Engine 8.1: EAN 13 supplemental), 2D PDF417
- **"Fast mode" options for ICR ,OCR, barcode recognition**
- **Data extraction from fields with various borders and frames**, combo-box, underlined fields, boxes, etc. and fields where the data are not fit into field border.
- **Despeckling of an image in individual blocks (or zones)**, with the ability to specify the size of black or white dots
- **Definition of field content**, by setting alphabets, dictionaries, regular expressions, types of segmentations, handwriting styles, etc.
- **Detection of in-field spacing**, accurately detecting small spaces, allowing use of dictionaries which may contain word-combinations with spaces.
- **Intelligent processing of blocks with intersecting parts and lines**, recognising text words and symbols which are completely located within block borders, avoiding the time spent for non-relevant text block information.
- **Voting API**, with word-level and character-level hypotheses for following voting scenarios
- **"On-the-fly" Recognition Tuning**, allowing integrators to influence the hypothesis choice by inserting additional ranking criteria during the recognition process.

## Data Capture (Fixed and Semi-structured Document Processing)

Fixed (templated) form and semi-structured document processing with support for ABBYY FormReader and FlexiCapture technologies.

- **Data extraction from machine-readable forms**, for export in databases and content management systems
- **Fixed (templated) form matching**, with automation identification of the templates
- **Semi-structured document (flexible form) template matching**, using ABBYY FlexiCapture technology. Reads documents with similar data by different layouts such as: invoices, shipping forms, templated based contracts, etc. Support for table elements, multilanguage selection, and special element types such as phone, currency, etc.



Using templates created with FlexiCapture Studio, FineReader Engine can process documents with layouts which vary from page to page

## ABBYY Licencing Policy

ABBYY FineReader Engine is sold via a flexible, modular licencing policy that allows developers to select the best combination of tools and pricing options for their project. Licencing is offered as:

### ■ Developer Licences

Providing right to develop and test applications integration FineReader technology. Special licences required for work with Asian Language recognition and fixed and flexible form matching via ABBYY FlexiCapture and FormReader technologies.

### ■ Runtime Licences

Grants right to distribute applications with FineReader Engine functions incorporated. RTLs differ by functionality, pages processed per month, and network support (Network Runtime License). The Professional Runtime Licence provides access to core recognition technologies. Additional RTLs oriented to specialised functions include the Barcode Runtime Licence and FineReader XIX Runtime Licence.

### ■ Add-on Modules for Runtime Licences

RTLs can be enhanced by adding one or more of the following functionality offered as add-on modules: PDF opening, PDF export, Word XML, Extended Character Information, CJK (Chinese, Japanese, Korean) OCR, 2D barcode recognition, ICR, IMR, document analysis for invoices, fixed form matching via ABBYY FormReader, semi-structured document matching via ABBYY FlexiCapture technology.

## Features Summary

### Input from Multiple Sources:

- Scanned images via TWAIN interface
- Digital camera
- Memory
- Image files including: BMP, PCX, DCX, JPEG, JPEG 2000, PNG, TIFF (including variable compression), DjVu, GIF
- Colour, greyscale, and black and white images
- PDF

### Image Preprocessing:

- Built-in adaptive binarisation and texture filtering
- Preprocessing for digital camera images
- Straighten lines function
- Automatic Image deskew
- Skew calculation based on information of reference blocks
- Rotate, mirror, and invert functions
- Image despeckling, scaling, and dipping
- Dual-page splitting
- Preview creation
- Text colour or background manipulation inside rectangles

### Page Analysis

- Auto-detection of page orientation
- Auto-detection of blocks, tables, barcodes, and pictures
- Auto-detection of vertical text in table cells
- Manual block manipulation (adding, removing, editing)
- Document Analysis:
  - Automatic page analysis (3 modes): normal, full text indexing, invoice pre-processing,
  - Manual document layout creation/editing

### Recognition

- **OCR:** Multilingual OCR and 191 languages (including Latin, Greek, Cyrillic alphabets, Chinese, Japanese and Korean), 41 with dictionary support; Text types: Normal, Matrix, Typewriter, OCR-A, OCR-B, CMC7, MICR, Fraktur/Gothic, mixed text type processing with autodetection on a word-level
- **ICR:** on digits, digits combined with letters of one language, and digits combined with letters of several languages, even if fields contain both upper and lower case letters. Separates field content from borders and frames. 91 languages, 22 with morphology, field-special dictionaries; 22 handwritten styles including English, American, German, French and Russian
- **Barcodes:** including processing of barcodes that are damaged or printed at any angle and fast barcode extraction, more than 16 most popular 1D industrial types, 2D PDF 417
- **Check mark (OMR):** simple, grouped, model check marks, marks with "corrections" made by hand
- **"Fast mode" recognition options** for OCR, ICR, barcode recognition
- **Core parameters tuning for customising processing speed** (switching on/off certain pre-processing, document analysis and recognition algorithms)
- **Pattern training**
- **User language support**

### Export/Output of Recognised Text

- Multiple levels of text format retention (simple text to complete layout retention)
- Access to detailed information about each recognised character
- Recognised text export into various formats: PDF, incl. PDF/A, RTF, TXT, XML, HTML, PPT, XLS, CSV, DBF
- Full picture and text colour retention
- Post-editing, post-formatting functions

© 2007 ABBYY Software. All rights reserved. © 1987-2003 Adobe Systems Incorporated. Adobe® PDF Library is licensed from Adobe Systems Incorporated. Fonts Newton, Pragmatica, Courier © 2001 Paratype, Inc. Font OCR-v-GOST © 2003 Paratype, Inc. ABBYY, the ABBYY Logo are registered trademarks or trademarks of ABBYY Software Ltd. Adobe, the Adobe Logo, the Adobe PDF Logo and Adobe PDF Library are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States and/or other countries. Microsoft, Excel, PowerPoint, Outlook, Windows are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. Unicode is a trademark of Unicode, Inc. All other trademarks are the property of their respective owners.