ABBYY Recognition Server
Product Information

Robust Document Capture and PDF-Conversion

ABBYY Recognition Server is a powerful server-based OCR solution for automated document capture and PDF conversion. It allows organisations and scanning service providers to implement cost-efficient processes for converting paper and image documents into electronic files suitable for long-term digital archiving and full-text search.

ABBYY Recognition Server automatically acquires document images from scanners, file, fax and e-mail servers, as well as Microsoft® SharePoint® libraries, performs optical character recognition to retrieve full-text information and offers the possibility to add metadata. The results are delivered directly to network folders, SharePoint libraries or other storage and management systems as MRC-compressed searchable PDF or PDF/A files, XML data, Microsoft Word and Excel® files or plain text.

This highly scalable solution allows you to quickly convert large quantities of documents in a short time. Its quick deployment, easy administration and automated work routines make ABBYY Recognition Server an investment that delivers fast returns.

How can you benefit from ABBYY Recognition Server?

The possibility to convert business documents into digital files in an automated way supports a variety of business processes, for example:

Converting Entire Archives to Searchable PDF/A and PDF Format

ABBYY Recognition Server automatically converts extensive collections of paper documents, scanned document images and complete books into PDF or PDF/A files that can be electronically archived, easily detected via keywords by e-discovery and enterprise search systems, or remotely accessed by employees or clients. The solution is highly scalable and can process large amounts of documents within tight timeframes.

Deployment of a Document Conversion Service

ABBYY Recognition Server allows implementing a centralized OCR service instead of installing OCR software on many individual workstations. Any employee within an organisation or a workgroup can convert scanned documents to Microsoft Word or searchable PDF files, reaching the service directly at the scanning point or from any location via email or FTP folder. The service can be deployed locally for own employees or in a hosted environment for external clients.

Creating Full-text Searchable SharePoint Libraries

Documents stored in Microsoft SharePoint, such as TIFFs created by fax servers and image PDFs created by scanners, remain "invisible" for the search engine and can't be detected. ABBYY Recognition Server can retrieve such files, convert them into a searchable format, such as PDF, and store them in the same location so they can be included in the search engine's index. If scanned PDFs exist among stored PDF files, the application can smartly detect them. It will then add a new text layer, turning them into searchable PDFs instantly.

ABBYY Recognition Server can crawl SharePoint libraries and network areas on a continuous basis and automatically convert all newly added image files. Should an existing TIFF collection need to be preserved in its original format, ABBYY Recognition Server can generate searchable text for those images and deliver it as an XML file to the Microsoft Search Engine or the Google Search Appliance leaving the original TIFFs in place.

Automated Document and PDF Conversion

Feature Overview

ABBYY Recognition Server converts documents automatically, with minimum user intervention. It runs in the background and independently performs all document processing steps - round the clock or at pre-defined times:

Step 1: Scanning and Document Input

Scanning

The application offers an easy to use Scanning Station interface that supports scanning of documents in batches. It provides tools for document quality improvements, such as image preview and enhancement, manual redaction, and many others. Scripting commands can be used, for example, to auto-split large pages or re-order pages after duplex scanning.

Document Import

Previously scanned document images can be automatically retrieved from document libraries or received per e-mail. The imported document images will be processed with corresponding priorities and according to available computing resources.

Scanning via TWAIN, WIA, ISIS
Integrates with all network scanners and MFPs.

Hot Folder Watching (FTP or Local Network)
Automatically processes files arriving in defined folders.

Crawling of Network Shares and SharePoint Libraries
Detects newly added image files and converts them into a searchable format.

Input via E-mail (Exchange, POP3)
Integrates with fax and e-mail servers and processes image attachments.

ADVANCED PDF PROCESSING

• Creates MRC-compressed PDF and PDF/A files that significantly reduce size of color documents.

• Supports encryption: Limits opening and printing of the created PDF documents.

• Detects scanned PDFs and PDFs with insufficient text quality and adds a new text layer to the original file.

• Retains original image, bookmarks, and attachments when inserting a new text layer into original PDF.

• Digitally created PDFs with a good text layer can be moved directly to the new location.

• Support for long term document archiving standards: PDF/A-1a, 1b, PDF/A-2a, 2b, 2u

• Creates PDFs optimised for Internet download.

Step 2: Processing of Documents

Document Recognition/OCR

The optical character recognition process runs automatically on a dedicated workstation – the Processing Station. Using ABBYY's award-winning OCR technology the system supports a broad range of functions to increase the recognition accuracy, including:

• **Image pre-processing** (for example split dual pages for book scans or clear background noise)

• **Print type definition** (chose between normal text, typewriter, dot-matrix, OCR-A, OCR-B, and MICR E13b)

• **Language definition** (more than 190 languages and historic texts in old fonts)

Depending on the document's quality and its structure, the processing mode can be set on either 'precision' or 'speed'. To increase processing speed significantly, for example to process many documents within a tight deadline, additional Processing Stations or a higher number of CPU-cores can be added.

Verification (optional)

In some cases, for example when digitising books, verification of the recognition results is necessary. The integrated Verification Station interface offers the possibility to correct the results either on all documents or only on documents that did not reach a predefined recognition accuracy threshold.

Indexing (optional)

If required, document indexing can be done either manually using the Indexing Station interface or automatically by a script. Lists of index field values can be imported and synchronised with third party systems.

Scheduled Processing
Different kinds of documents can be processed at different times according to a schedule.

24/7 Fail-safe Processing
Multiple Processing Stations and cluster deployment can be used to distribute the workload dynamically and assure reliable processing.

Barcode Recognition
Values of most popular 1D and 2D barcodes including 2D Aztec, Data Matrix, and QR Code barcodes can be extracted.

Recognition of Historical Texts in Old Fonts
Support for black letter, Schwabacher and most other Gothic fonts in English, German, French, Italian and Spanish.

Automated Document and PDF Conversion

Feature Overview - continued

Step 3: Assembly and Export

After the recognition stage, ABBYY Recognition Server assembles the processed pages into individual documents. The documents can be separated using blank sheets or barcode pages as separators or by a fixed number of pages per document. Separation can also be done according to a scripted rule.

Assembled documents in the required formats are delivered to predefined output locations such as network folders, SharePoint document libraries, and e-mail addresses. They can as well be handled over to other applications connected via the API.

Additionally, scripts can be applied for intelligent routing and delivery of documents to Enterprise Content Management systems based on document properties and attributes. ABBYY Recognition Server supports a variety of output formats and allows creating several output files at the same time.

To turn digital archives into fully searchable electronic document archives the application can crawl individual libraries, detect not searchable image-based documents and convert them into searchable formats. Documents such as Microsoft Word files, PowerPoint® presentations or Excel spreadsheets, which don't require any processing, can be moved into the output library to the same position. This way any document library can be turned into fully searchable electronic library.

Multiple Output formats
Variety of formats, including searchable PDF and PDF/A (MRC-compressed), XML, RTF, Microsoft Office and others

Publishing to Network Folders
The original folder structure is automatically mirrored. The name of output files can be flexibly defined using a barcode, the document type, etc.

Sending by E-mail
Converted documents can be delivered back to the sender or to a list of specified recipients.

Publishing to SharePoint
Results can be automatically uploaded to SharePoint libraries. Scanned PDFs stored within SharePoint can be enhanced with a text layer and saved under a new version number.

ABBYY 3A
Asia, Baltic, Middle East, South America, Africa
P.O. Box #32, Moscow, 127273, Russia
Tel +7 495 7833700
Fax +7 495 7832663
sales_3A@abbyy.com
www.ABBYY.com

ABBYY®

SPECIFICATIONS

General System Requirements
- **PC** with Intel® Core™2 /2 Quad /Pentium® /Celeron® /Xeon™, AMD K6/ Turion™/ Athlon™/ Duron™/ Sempron™ processor with min. 2 GHz
- **Operating system:** Microsoft® Windows® 8, Windows 7, Windows Vista®, Windows Server® 2012, 2008
  Memory (RAM):
  Server Manager: 1 GB
  Scanning Station: 1 GB
  Processing Station: 512 MB plus 300 MB for each recognition process
  Indexing Station: 768 MB
  Verification Station: 1024 MB
- **Hard Disk Space:**
  Server Manager: 20 MB for installation plus 1 GB for program operation
  Scanning Station: 1 GB
  Processing Station: 600 MB for installation plus 1 GB for program operation
  Indexing Station: 500 MB for installation plus 1 GB for program operation
  Verification Station: 700 MB for installation plus 700 MB for program operation.
  Requirements for program operation depend on complexity, quality, and number of images. System requirements may vary based on server component or additional module used. Contact ABBYY for more detailed specifications.
- Microsoft .NET Framework 3.5 or later for saving files to Microsoft SharePoint Server
- Microsoft Outlook® 2000 or later for processing e-mail messages
- Microsoft IIS 5.1 of later for Web API
- Scanner supporting TWAIN, WIA or ISIS

User Interface Languages*
English, French, German, Italian, Spanish, Russian, Portuguese (Brazilian), Czech, Hungarian, Polish, Chinese (Traditional and Simplified).
*Release 1 contains only Russian and English user interface.

Input Formats
- BMP, PCX, DCX, GIF, TIFF / Multipage TIFF, WDP, WMP
- JPEG, JPEG 2000, JBIG2, PNG, RLE
- PDF (up through version 1.7), DjVu, JPX

OCR Languages
More than 190 languages

Print Types
Normal, Fax (mode for low resolution texts), Typewriter, Dot matrix printer, OCR-A, OCR-B, MICR (E13B), Gothic

Barcode types
1D: Check Code 39, Check Interleaved 25, Code 128, Code 39, EAN 13, EAN 8, Interleaved 25, CODABAR (without checksum), UCC Code 128, Code 2 of 5 (Industrial, IATA, Matrix), Code 93, UPC-A, UPC-E, Patch Code and Postnet
2D: PDF 417, Aztec, Data Matrix, QR Code

Output Formats
Editable Formats
- RTF, TXT, HTML, CSV, EPUB
- DOC, DOCX, XLS, XLSX
- XML, Alto XML, FineReader internal format
Searchable Formats
PDF (up through version 1.7); PDF/A
Image Formats
- Image-only PDF, PNG, JBIG2
- JPEG, JPEG 2000, TIFF

Integration and Customisation Options
XML Tickets, COM-based API and Web Service API, Scripting in VBScript and JScript