

# ABBYY Recognition Server™

## Server-Based OCR and PDF Conversion Solution

ABBYY Recognition Server is a robust, server-based solution for automating document recognition and PDF conversion processes. It is a scalable, reliable and rapidly deployable solution for high-performance delivery of optical character recognition (OCR) functions in a service-oriented environment. OCR functions can be easily used together with existing infrastructures and third-party applications.

### OCR via a Service-Based Architecture

ABBYY Recognition Server places all of the recognition processing functions on a server. As a result, customers benefit from the following key advantages associated with service-based architectures:

#### ■ Automated OCR processing with centralized management

Administrators can pre-define settings and control functions across the enterprise. A remote management console serves as a central point for defining processing parameters and managing recognition stations so that they can create specific “workflows” for a particular workgroup or project. Corporate users can access powerful recognition functions without the need for individual training.

#### ■ Independent, server-based processing

Recognition and conversion are performed remotely on the server as a background process, working independently without affecting other tasks running on client workstations. Processing can take place unattended and “instantly,” upon the placement of a document in a certain “watched folder,” or at scheduled times.

#### ■ High-volume processing, extendable scalability

Recognition Server can handle any volume of documents because it is not limited by the constraints of a desktop workstation. Concurrent processing with automatic job distribution maximizes throughput. Processing power is easily extended by simply adding additional processing stations and CPUs, because the program’s Server Manager can easily manage clusters of recognition stations.

#### ■ Fast deployment, simple integration, and interoperability

Flexible enough to respond to rapidly changing business needs, Recognition Server can be deployed as a complete stand-alone solution or work with an external application. It is easy to deploy Recognition Server because it offers:

- **Ready-to-use functions.** A standard set of ready-to-use functions for key processes (e.g. converting images to searchable PDFs) requires minimal pre-configuration to get up and running.
- **Open API.** Companies can add powerful OCR without disrupting their existing document management process.
- **XML-ticket support.** Enables instant management of processing from the client side or third-party applications.

#### ■ Fault tolerance

Includes built-in functionality to ensure ongoing stability of the system and thus protect data. Functions include server pinging to ensure proper functioning, server/job logging, automatic restart, and job re-routing in case of a system failure.

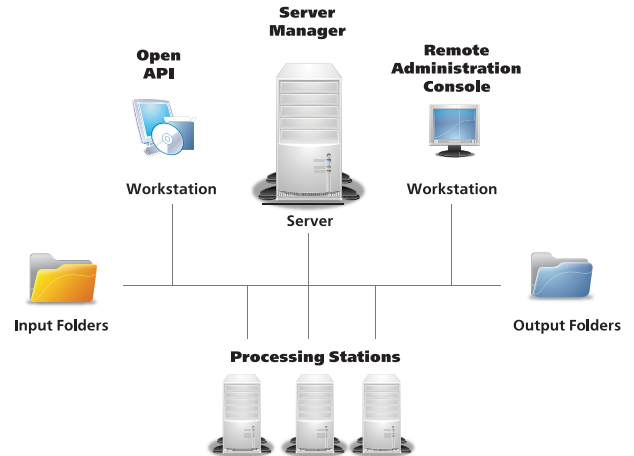


*Based on ABBYY's award-winning document recognition platform, ABBYY Recognition Server is a turnkey solution for service-based document recognition and PDF conversion. It delivers highly accurate conversion of PDF files, images and scanned documents into text for the purpose of categorizing, searching and export into third-party systems for document storage and management.*

# How ABBYY Recognition Server Works

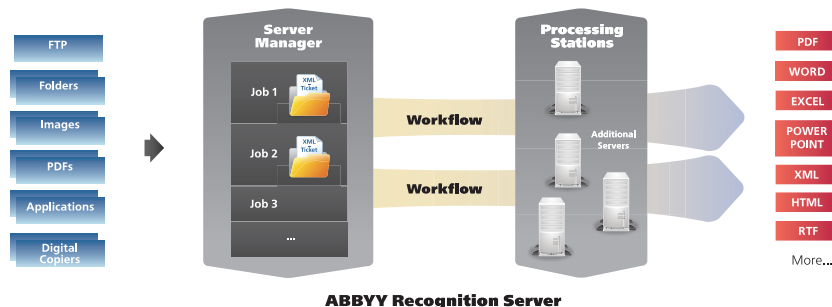
**ABBYY Recognition Server is comprised of the following components:**

- **Server Manager.** Manages all requests, processing options and services and distributes various recognition and conversion jobs among processing stations.
- **Processing Station.** Performs OCR and document conversion. Receives jobs from the Server Manager. The system is easily scalable by adding additional processing stations.
- **Remote Administration Console.** Based on the Microsoft Management Console (MMC), provides a common administration interface for configuration and monitoring of the system. It can be installed with the Server Manager or on a separate workstation.
- **Open API.** COM-based Open API enables management of Recognition Server functions from within external applications (e.g. ECM systems), linking the product to third-party workflows. Can also be used to override processing settings for the purpose of automating and customizing jobs. VARs or third-party developers can use the API to integrate Recognition Server with their own client applications.



**Recognition Server manages tasks via two main elements - workflows and jobs:**

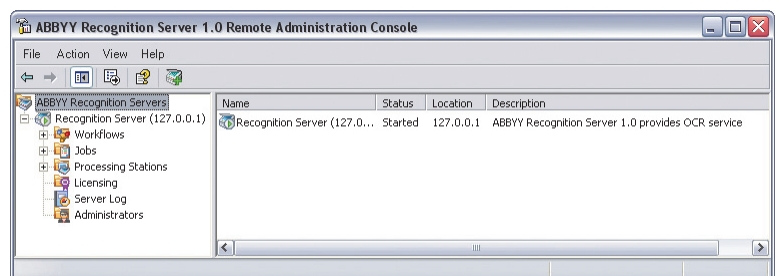
- **Workflow.** The automation of a business process in which information and tasks are passed from one resource to another for action according to a set of procedural rules. Workflow parameters include such items as scheduling, processing criteria (e.g. recognition scenario, language), output formats, and error handling.
- **Job.** A unit of processing (e.g. a set of documents and the processing parameters for this set). Jobs can contain one or several documents and inherit the processing parameters of the workflows in which they were created. Jobs are also accompanied by **XML-tickets** to manage parameters which may be used to extend or override the workflow settings.



**A Recognition Server administrator defines all Recognition Server parameters via the remote administration console.**

From this central point, the administrator controls the following functions:

- **Workflow properties**
- **Job management**
- **Processing stations**
- **Server log**
- **Administrators**



CONT

WORKFLOW NUMBER

# Key Functionality

Recognition Server offers powerful recognition and conversion functions which can all be managed centrally. Organizations may control and customize the workflows and OCR processing and monitor the system's productivity with the following functions:

## Input

**Document Input Folders.** Recognition Server offers two options for adding documents to a workflow. Documents can be transferred from a client computer or directly from a networked scanner or multifunctional device (MFP) to either a **shared folder** or **FTP folder**.

**Workflow scheduling and "watch folders."** Processing can be performed according to a pre-determined schedule, or "instantly" on certain "watched" folders as soon as new images are added. There are options for selecting time, start and end date, and ongoing scheduled processing such as weekly or monthly tasks, so that jobs can be set to run overnight, or during off-peak hours for maximum efficiency.

**Workflow priority.** Administrators can assign priorities to their processing profile.

**Image input formats.** Supports popular formats such as BMP, PCX, TIF or PDF\* (including image-only PDF).

## Processing

**Pre-defined processing scenarios.** Administrators may select the conversion or recognition method optimized for a specific task:

- Full-page document conversion (with formatting and layout retained)
- Full-text indexing
- Invoice pre-processing
- Extract barcodes only

## Recognition

- **Print type.** Normal text, typewriter, dot-matrix, OCR-A, OCR-B, and MICR (E13b).
- **Language.** 184 languages, multilingual documents.
- **Special fonts.** Fraktur, Schwabacher and majority of Gothic fonts printed between 1700 and 1937 in English, French, German, Italian and Spanish.\*
- **Barcodes.** Recognizes most popular 1D and 2D barcodes at any angle on a document.
- **Speed control.** Options for both thorough and fast-mode recognition.

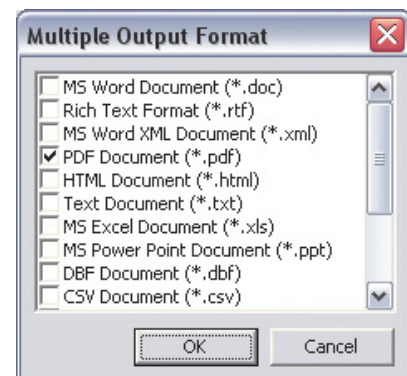
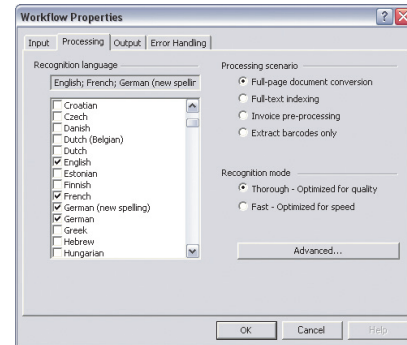
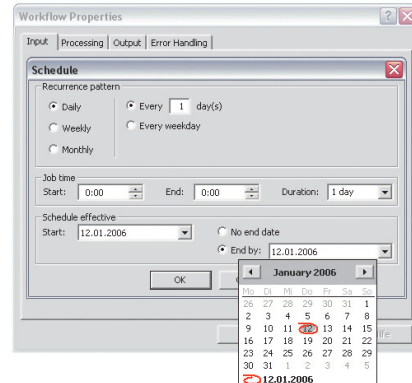
**Advanced processing and image pre-processing.** Selection of automatic functions such as split dual pages (for book scans), convert color and grey to black and white, clear background noise, and image deskew and despeckle.

## Output

**Multiple document output formats.** Includes DOC, RTF, XML\*, DBF, XLS, HTML, PDF, etc. Can be set up to create multiple document output formats for a single process (e.g. convert image-only PDFs to searchable PDF and XML files in a single pass). Supports searchable PDF files and linearized, tagged, and secure PDF formats. Supports native PDF encryption – 40-bit and 128-bit AES.

## Error Handling

- **Quality control.** Ability to set a quality threshold based on the confidence level and percentage of uncertain characters per page.
- **Exception folder.** Separates pages with poor results into a unique folder.
- **Job cancellation.** Sets a timeout for automatic cancellation of a job.



# When to Use ABBYY Recognition Server

Recognition Server's server-based architecture is well suited for a variety of applications:

- **Document recognition running as an unattended operation.**  
Scheduling functions and "watched folders" enable processing to run simply at the insertion of a document into a file folder, or at night, on weekends, during off-peak hours, etc.
- **Distributed scanning with easily managed centralised processing.**  
Well suited for use in a multi-user, multi-location corporate environment. An insurance agent, for example, may have documents scanned at each local office which then submit the documents to the server for central processing.
- **Mid- to high-volume processing.**  
Multiple processing queues and easy scalability through the addition of multiple processing stations.
- **Distributed scanning with easily managed centralised processing.**  
Well suited for use in a multi-user, multi-location corporate environment. An insurance agent, for example, may have documents scanned at each local office which then submit the documents to the server for central processing.
- **Business process outsourcing (BPO) applications.**  
Ideal for service providers because it is easy to configure settings for a variety of different jobs.
- **Corporate environments requiring centralised processing suited for multiple departments or functions.**  
Set up different workflows to accommodate different processing and routing needs based on document types. For example: CVs from the HR department for indexing or contracts from the sales department for conversion to PDF for email attachments.



## System Requirements\*\*

- PC with Intel® Pentium® or compatible processor with 500 MHz or higher, running operating systems Microsoft® Windows® Server 2003, Windows XP or Windows 2000.
- 128 MB of RAM, and additional 100MB for each additional recognition process.
- 350 MB of available hard disk space for the installation of server, recognition stations, and program operation.
- Administrator account using Recognition Server may require read/write permission to certain registry branches (see manual for details).

## Pricing and Availability

Recognition Server is sold as a standard license package with additional add-on functionality. This modular licensing scheme flexibly enables customers to add only the functionality which is required.

A basic Recognition Server license includes core recognition functionality and support for a variety of export formats including PDF.

Available add-on modules include:

- **PDF opening**
- **FineReader XIX (Fraktur/Black Letter font recognition)**
- **Chinese, Japanese and Korean recognition**
- **Export to XML**

Pricing is based on a "cost per CPU" pricing model. License contracts run in 1-year time periods and are renewable each year with the purchase of extended software maintenance and upgrade assurance (SMUA) contracts.

\* Available as add-on modules.

\*\* System requirements for installing the Server Manager, recognition station and management consoles on separate computers may differ.

## Specifications

### Image Input Formats:

- BMP
- PCX
- JPEG
- JPEG 2002 part 1
- PNG
- TIFF
- PDF (up to PDF 1.6)\*

### Output Formats:

- DOC (Microsoft Word)
- RTF
- XML (Microsoft and Native)\*
- PDF (version 1.6)
- HTML
- PPT (Microsoft PowerPoint®)
- CSV (comma-separated values)
- TXT
- XLS
- DBF



©2006 ABBYY USA Software House. All rights reserved.  
©1987-2003 Adobe Systems Incorporated. Adobe PDF Library is licensed from Adobe Systems Incorporated. ABBYY and the ABBYY logo are registered 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, Windows are registered trademarks of Microsoft Corporation in the United States and/or other countries. Fonts Newton, Pragmatica, Courier ©2001 ParaType, Inc. Font OCR-v-GOST ©2003 ParaType, Inc. All other trademarks are the property of their respective owners. Information contained in this document is subject to change without notice.

ABBYY USA Software House, 47221 Fremont Boulevard, Fremont, California 94538, United States.  
sales@abbyyusa.com. www.ABBYY.com.