

What's New

ABBYY® Recognition Server™

ABBYY Recognition Server 3.0 What's New

ABBYY Recognition Server 3.0 is a powerful and user-friendly document capture system that allows organizations to quickly establish a cost-efficient process of converting paper to searchable and editable electronic documents. Recognition Server 3.0 provides functionality for scanning paper documents, optical character recognition (OCR), extracting metadata and delivering the content into e-archives, enterprise content management and search systems.

No matter what the project time constraints, Recognition Server's scalable architecture can quickly process huge volumes of documents. Simple to deploy and use, Recognition Server 3.0 ensures the work is done on schedule with minimal effort.

With the addition of full text conversion, rubber-band indexing and solutions for enterprise search systems integration, Recognition Server 3.0 has evolved from providing document conversion into a robust, high-volume document capture solution.

Key New Features

1 Scanning Station

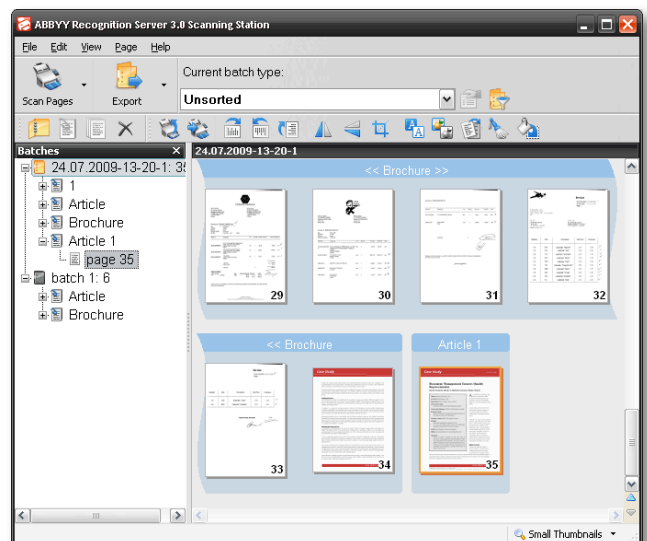
The new Scanning Station allows documents to be scanned directly into workflow via TWAIN, WIA and ISIS drivers. In addition to monitoring FTP and hot folders for input documents, now document conversion can start right from the scanner.

The Scanning Station offers an intuitive interface for high-speed production scanning allowing the operator to:

- Start/stop scanning and adjust scanning options
- Control the quality of scanned images "on the fly"
- Assemble documents for further OCR and indexing

The Scanning Station provides the following functionality:

- TWAIN and ISIS scanning
- Image import
- Image enhancement:
 - Rotate clockwise, counter-clockwise, upside down and flip
 - De-skew and de-speckle
 - Convert to black and white
 - Invert colors
 - Crop
 - Redact
- Document separation by barcodes, blank pages and fixed number of pages
- Export to local folder, FTP folder or Recognition Server



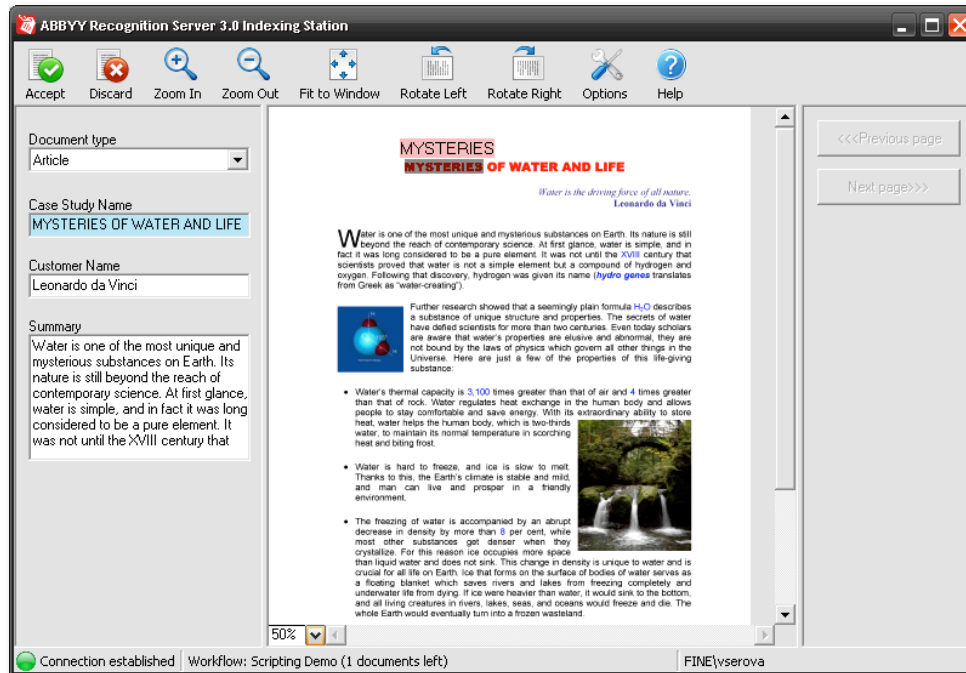
What's New

2 Indexing Station

Simple indexing tasks now can be performed with a few mouse clicks. The point-and-click indexing feature eliminates the need to retype the text; because the data has already been pre-recognized, the operator just needs to select which data should be used for indexing.

The Indexing Station allows the operator to:

- Select the document type from the drop-down list
- Fill in the indexing fields using the recognized text (“point-and-click” indexing) or manually if required
- Send the document to export or simply discard it



The Indexing Station provides the following functionality:

- Allows operator to select from administrator-defined types of documents
- “Point-and-click” indexing: click on the words and outline paragraphs to transfer them to attribute fields
- Manual indexing (key from image)
- Automatically highlights words that match the regular expression within the document
- Assigns operator for indexing tasks

3 Scripting

New scripting capabilities allow Recognition Server 3.0 to be customized and seamlessly integrated into existing processes. New features include scripting for **document separation**, **indexing** and **export handling**. These scripts allow administrators to create more flexible document separation rules, automate document type detection and indexing, and simplify exporting into back-end systems.

Scripts can be used to accomplish the following:

- **Assemble documents** - Scan multiple employee documents in a batch and apply employee number as an attribute to each document. Use cover sheet with employee # on top of batch, and use blank page separators for each document in the batch. Script applies employee # to each document on export to ERP system or SharePoint.
- **Customize file naming** - Output result files can be customized if the default needs to change.
- **Route documents** - Export output documents to folders created based on document type, index fields and recognition statistics for easier search and access.

What's New

- **Index documents** - If the desired index fields are always located at the same area on the front page of the document, the indexing process can be automated using scripts.
- **Send by e-mail** - Send documents to different recipients based on the index data.
- **Error handling** - Send custom notification emails to the administrator via SMTP with authentication.

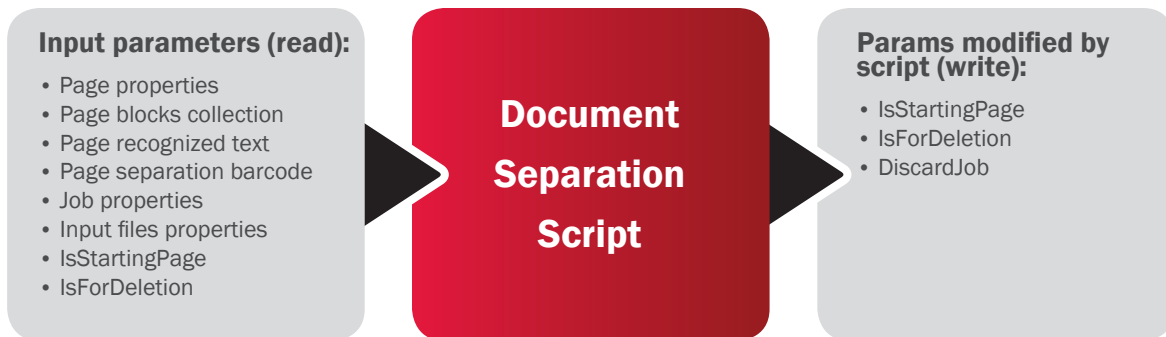
The following describes the functionality of the three scripts provided with Recognition Server 3.0.

A. Document Separation Script

This is used to customize document separation rules, which are in addition to built-in document separation methods. This script is executed on every page after OCR and verification.

The document separation script can:

- Set the beginning of a new document based on the page content and properties
- Mark a page for deletion (e.g., a separation page or a blank page)
- Discard a job



B. Document Indexing Script

The Document Indexing Script is used to automate document type detection and indexing as an alternative or supplement to point-and-click indexing. Executed on every document after separation and before manual indexing.

The document separation script can:

- Set document type and attributes based on the document content and properties
- Decide whether or not the document should be routed to the indexing station



What's New

C. Export Handling Script

Export handling scripts are used to handle the resulting files and data delivered by Recognition Server 3.0 and to perform additional operations after processing, as shown below:

Export Handling Script

For Output Documents	For Failed Jobs
Executed on every document published to Output Folder	Executed on every failed job
Handles output files and data (e.g., sends them to an ECM system)	Handles images that failed to be processed

Input parameters (read):

- Output files
- Document type and attributes
- OCR statics
- Job properties
- Input files properties
- XML result

**Export Handling
Scrip output docs**

Input parameters (read):

- Job properties
- Input files properties
- Error message
- XML result

**Export Handling
Scrip failed jobs**

4 New Recognition Languages

The new Recognition Server 3.0 languages are Vietnamese with dictionary support and Chinese Simplified (PRC), Chinese Traditional (Taiwan), Japanese, Korean, Korean Hangul, Yiddish and Latvian Gothic without dictionary support.

5 New 2D Barcodes

In addition to PDF 417 Recognition Server 3.0 now also recognizes Aztec, Data Matrix and QR Codes.



Aztec



DataMatrix



QR Code

What's New

6 ADRT

ABBYY ADRT significantly improves the document layout retention when saving documents to DOC and RTF. The logical internal structure of an entire document is reproduced, including headers, footers, footnotes, page numbers, table of contents linked to the document sections, and notes to pictures and diagrams.

7 Connectors to Google® Search Appliance and Microsoft® Search

Recognition Server 3.0 expands the capabilities of Google® and Microsoft search systems adopted in the organization by offering connectors to the Google Search Appliance (GSA), Microsoft Office SharePoint® Server and Windows Search. Recognition Server 3.0 makes all documents easily searchable, since in addition to the conventional indexing of text-based documents, it enables the search system to index the full text content of faxes, scanned agreements, correspondence and any other TIFF, JPEG or PDF documents.

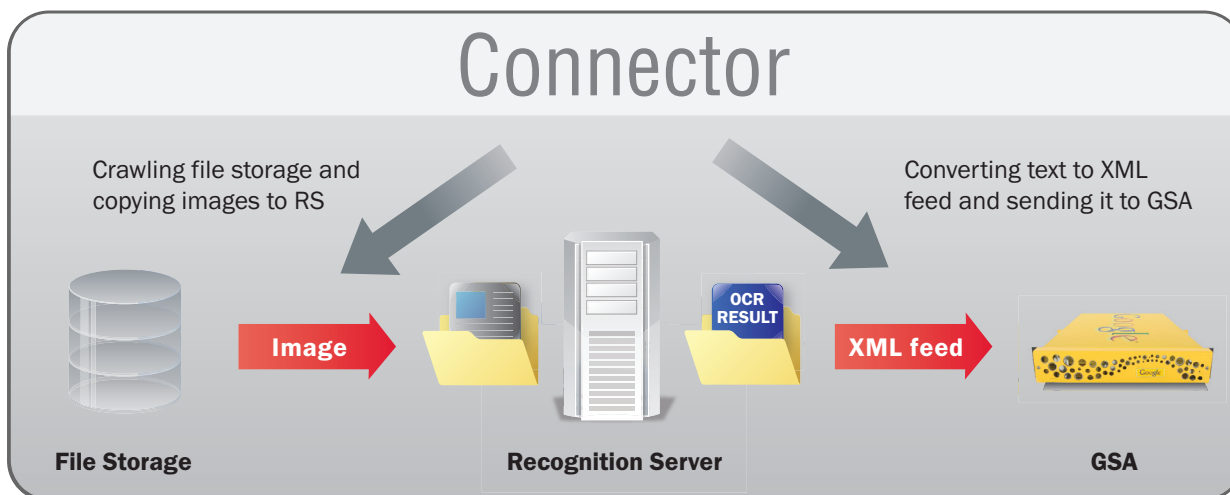
A. GSA Connector

GSA is an enterprise search system developed by Google. It is a hardware-software solution dedicated to indexing documents across a company's file servers, web-servers and other resources making them available for search through a Google-like interface.

GSA supports full-text indexing of searchable and editable file formats, such as HTML, DOC, XLS, TXT and text-based PDF, and on file and web servers as an "out of the box" feature. Recognition Server 3.0 offers a ready connector to GSA that enables the system to index the contents of image documents as well.

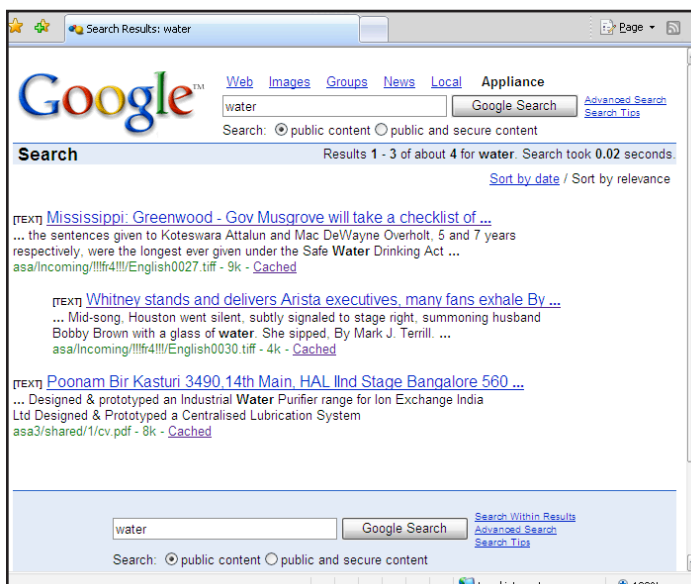
The GSA connector performs the following functions:

- Crawls specified network file shares for image files and sends them to Recognition Server for OCR
- Picks up the OCR results and submits the recognized document text to GSA as an XML feed, accompanied with a link to the original image



The text from the feed is included in the index by GSA. If the user types a search query that contains a word from the document's text, the image will be delivered in the search results.

What's New



B. Recognition Server 3.0 iFilter

Recognition Server 3.0 offers a connector to the Microsoft Office SharePoint Server 2007 search and Windows Search.

By default, Microsoft search engines only can index full contents of documents in text-based formats, such as HTML, DOC, XLS, etc. Indexing of files in any other formats can be implemented via "iFilters"—special plug-in modules developed by third parties (e.g., PDF iFilter, which can be downloaded from Adobe's® web site, enables SharePoint and Windows Search to index and search through PDFs with a text layer).

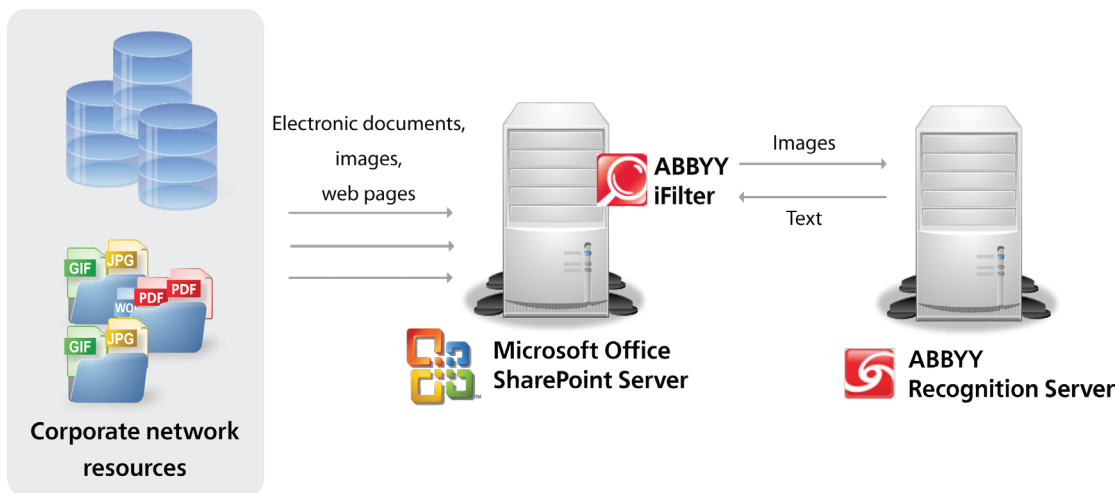
Recognition Server 3.0 includes an iFilter that enables SharePoint Server and Windows Search to index contents of scanned images and PDF documents in the following manner:

- It receives image files from the Microsoft search system crawler and sends them to Recognition Server 3.0 for OCR
- It picks up the OCR results and sends the full text content of the document back to the Microsoft search system engine for indexing

The text of the document is included in the index of the SharePoint Server (or Windows Desktop Search); the image then becomes discoverable through full-text search.

Integration with Microsoft Office SharePoint Server

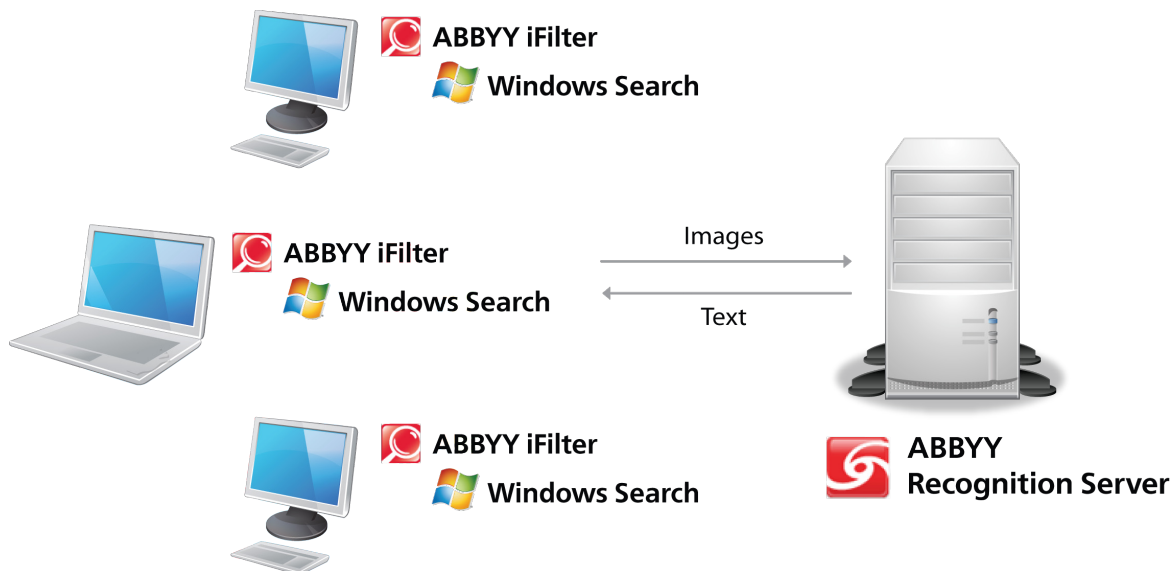
Recognition Server 3.0 may be installed on a separate server; the iFilter component must be installed on top of the SharePoint Server, as shown below.



What's New

Integration with Windows Search

Recognition Server 3.0 is installed on a separate server; the iFilter component is installed on desktop computers with Windows Search across the organization.



Additional New Features of Recognition Server 3.0

New image saving formats

- Gray or color TIFF Packbits compressed
- Black and white, gray, color TIFF LZW-compressed
- JBIG2

Patch code auto-detection – Patch codes can now be used as separation marks in a “Separate by barcodes” scenario.

Hebrew and Thai exported to all formats – Hebrew and Thai can now be exported to all formats supported by Recognition Server, including PDF and MS Word.

Improved accuracy on low quality documents like faxes - The new text type for faxes has been added.

Save to PDF 1.7 – Now administrators can distinctly specify the PDF version that Recognition Server 3.0 should generate, including PDF 1.7 files.