

Recognition Server 4 Release 6

Release Notes

Table of Contents

INTRODUCTION	3
About This Document	3
About the Current Release	3
Technical Information	3
Key Enhancements	3
NEW FEATURES AND IMPROVEMENTS	4
1. A new version 3.1 of ALTO XML	4
2. Ability to delete pages at Verification Station	4
3. Regular expression for separation barcode value	5
4. List of allowed barcode types	5
5. Fast opening of PDF files (invisible text layer).....	5
6. IFilter support for MS SharePoint 2016.....	6
7. Bug fixes	6
ABOUT THE PRODUCT	8
About the Product Version	8
Installing the Product Version	8
License Usage	8
History of Releases	8
Release 5 with Japanese Help Files Patch 1	8
Release 5 with Japanese Help Files	8
Release 5 Patch 1	8
Release 5.....	8
Release 4 for Symantec.....	8
Release 4.....	9
Release 3 with Japanese UI and Help.....	9
Release 3 Patch 2 (for customer)	9
Release 3 Patch 1 (for customer)	9
Release 3 with Japanese UI	9
Release 3.....	9
Release 2 Patch 2	9
Release 2 Patch 1	9
Release 2.....	9
Release 1 Multilingual	10
Release 1	10
Release 1 (specially for 3A).....	10
Arabic Edition	10

Introduction

About This Document

This document describes the improvements have been implemented in ABBYY Recognition Server 4 Release 6.

About the Current Release

ABBYY Recognition Server 4 Release 6 brings minor improvements and a number of bug fixes.

Technical Information

Part #: **1135/24**, build # **4.0.7.575**, OCR Technologies build # **13.0.35.70**, release date: **December 12, 2017**.

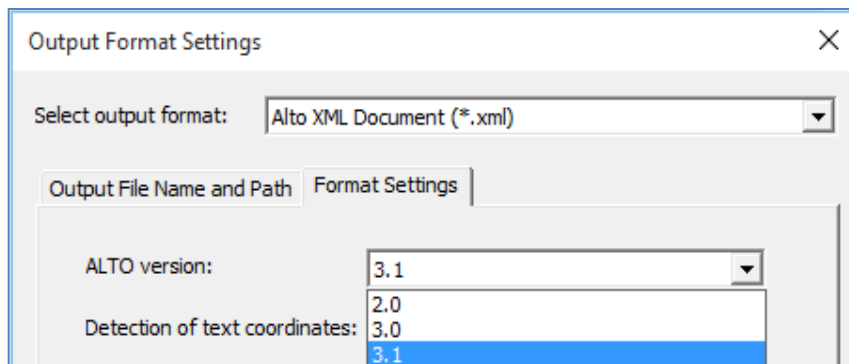
Key Enhancements

- Export to ALTO XML v.3.1
- Ability to delete pages at Verification Station
- Regular expression for separation barcode value
- Bug fixes

New Features and Improvements

1. A new version 3.1 of ALTO XML

Export to ALTO XML is extended with a new version of ALTO XML standard – version 3.1:
<http://www.loc.gov/standards/alto/ns-v3#> <http://www.loc.gov/alto/v3/alto-3-1.xsd>.



The basic support of ALTO XML 3.1 includes the ROTATION attribute in the TextBlock element for documents, which have the blocks of rotated text.

This attribute contains the angle of text rotation value: 90, -90, 180 (counting counterclockwise). The default value is "0". It is not specified if the text block is in normal orientation.

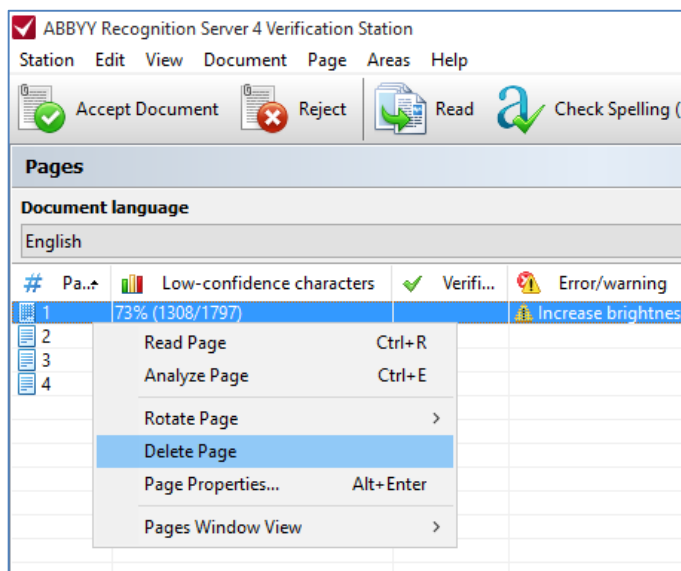
```
- <TextBlock ID="Page1_Block1" WIDTH="348" HEIGHT="47" HPOS="286" VPOS="824"  
  LANG="en-US" STYLEREFS="StyleId-B11A41B4-2512-4153-A42E-DA72E219CAE3-"  
  ROTATION="-90">
```

2. Ability to delete pages at Verification Station

Verification Station allows deletion of pages from the document.

This helps to modify the document structure in case the excess pages were scanned by mistake; the page occurred in the document due to the separation error and should be scanned as a part of another document, etc.

To remove a page, use **Delete page** command from the context menu of a particular page.



3. Regular expression for separation barcode value

When the separating of documents by barcodes is enabled, it is possible to specify the regular expression in the Configuration.xml.

It is valuable if there are several barcodes printed on the document pages. Some of barcodes are used for separation; some of them contain another values, for example encoded information from the document data.

The alphabet of regular expressions is described in the Help file (**Regular Expressions** article).

```
DocumentSeparationMethod="SeparateDocumentsByBarcodeOnTheFirstPage" DeleteBarcodePages="false"
BarcodeRegExp="1234567[0-9][0-9][0-9][0-9][0-9]" BarcodeType="Ean13">
```

By default, the parameter's value is empty: **BarcodeRegExp=""**.

Regular expressions for barcode values are also available via COM API of Recognition Server.

4. List of allowed barcode types

It is now possible to specify the list of barcode types that should be detected in the document.

By default, all supported barcode types are used during the document analysis. If a customer uses the specific barcode type, which has specific subtypes, it may lead to a wrong detection of the barcode type.

For example, Code39 is the main type, the rest of barcodes below are subtypes. When using the default settings, Code39 barcodes may be detected as of Code32 type and result in wrongly recognized value.

```
Code39
* CheckCode39
* Code39WithoutAsterisk
* Code39FullASCII
* Code32
```

In the workflow settings, it is possible to use only one allowed barcode type used for separation. A new list of barcode types allows using several barcode types during the document analysis and being more flexible when processing documents with multiple barcodes.

```
RecognitionMode="FullPage" VerificationMode="NoVerification"
<TextType>Normal</TextType>
<Language>English</Language>
<AllowedBarcodes>Code39</AllowedBarcodes>
<AllowedBarcodes>Datamatrix</AllowedBarcodes>
</RecognitionParams>
- <ExportParams IsExportSrcEdocAllowed="true" XMLResultPublish
DeleteBarcodePages="false" BarcodeType="Unknown" Docume
```

The list of allowed barcode types can be specified by editing the Configuration.xml file.

The complete list of barcode types allowed to be detected is made of the list of allowed barcodes specified in **AllowedBarcodes** parameters, plus the barcode type selected in the document separation settings (if applicable).

The list of allowed barcode types is also used re-recognizing a page at the Verification Station. In case of manual block editing, the operator can select any of the supported barcode types in the block properties.

5. Fast opening of PDF files (invisible text layer)

Thanks to the modified procedure of PDF file creation, PDF files produced by Recognition Server can now be opened, viewed, scrolled and scaled significantly faster. This is especially useful for PDF files made of the color images, construction drawings and large pages with many details (multiple tiny objects).

The modification is in saving the text layer as invisible. This helps to reduce the time required to display the file content. At the same time, it does not influence on copying the text or searching among the PDF file. All operations with PDF file work as usual.

Previously this feature was available via the parameter **FastPagePreview** of the Configuration.xml file of Recognition Server settings. To disable the feature of invisible text layer, change the **FastPagePreview** parameter's value to "False".

6. IFilter support for MS SharePoint 2016

It is possible to use IFilter component for indexing images stored in Microsoft SharePoint version 2016. This is supported in Recognition Server 4 installation wizard by default now.

(In the previous version, it was possible with an additional installation key only.)

7. Bug fixes

The current release of Recognition Server includes a number of bug fixes.

Description	SubSystem
Processing error: Internal program error: .\Src\AustraliaPostDecoder.cpp, 416	Barcodes
Processing error: Internal program error: .\Src\TextRecognizer.cpp, 487	Barcodes
Processing of the attached documents is failed with division by zero error.	Barcodes
Code-39 barcode is recognized as Code-32 on the attached documents.	Barcodes
The certain document with a table on Italian: a piece of the table is not recognized, the numbers in the cells are replaced by #.	Document Analysis
Processing error: Internal program error: .\Src\DocumentModelGenerator.cpp, 125	Document Analysis
The coordinates of internal block exceed the external block's coordinates.	Export
Processing error: An error occurred while exporting the result: Not enough memory!, export profile #1.	Export
The page image is lost after injecting the text layer into PDF	Export/PDF
The page images are lost (completely blank or completely black) after injecting the text layer into PDF/A	Export/PDF
Adobe Reader shows the warning "cannot extract the embedded font "Arial-BoldItalicMT" some characters may not display or print correctly", when opening the exported PDF	Export/PDF
Adobe Reader shows the warning "cannot extract the embedded font", when opening the exported PDF file	Export/PDF
Opening an output file in Adobe Reader 9.4.0/Acrobat 9 Pro fails with an error: An error exists on this page. Acrobat may not display the page correctly.	Export/PDF
Modify text layer only. Some metadata are lost after recognition.	Export/PDF
Expected an array object when splitting the pages	Export/PDF
It is necessary to describe that TextExtractionMode property is the same as Extract text from pictures feature	Help
Partly not translated text in the Admin Guide Eng, page 43	Help
Help file includes articles describing the missing functionality of custom languages creation.	Help
Empty page in a Help file: User role dialog description.	Help
Processing error: Internal program error: e:\teamcity.recognitionserver.4.0\technology\trunk\0\image\libraries\toolset\src\rlfrombitonalbitmapstreamfetcher.cpp, 30	Image
The spacebar doesn't fit any "." character at the beginning or end of the regular expression for validation	Indexing
IFilter doesn't work for SP 2016 by default	Installation
Processing through the Office is not localized	Resources
German. The administration console, Jobs view, Status column. Task status "N% getan" should be "N% erledigt" in German.	Resources
Processing error: Internal program error: Division by zero when processing attached file	Server
Table headers text could not be recognized in the attached document.	Server
Processing hangs for a certain file.	Server

Verification Station. There are settings Save Selected Pages As and Save Selected Pages To in the right-click menu. These settings have the drop down menu with always disable options.	Server
XmlResult: IsFailed parameter is not changed from False to True after processing of the erroneous document.	Server
There is an error: \DocumentAnalysis.GradientImages.aux contains an invalid path, when processing a document with a long file name	Server
The workflow loads the CPU after publishing the first file.	Server
Processing error: Internal program error: Rational overflow.	Server
Processing error: Internal program error: .\src\JobManager.cpp 925. when using a Scanning Station	Server
Processing stations do not use parameters specified for CPUs	Server
An undefined message ERR_AOO_CONNECTOR_NOT_REGISTERED	Server
If more than 1000 documents are queued for verification, then Verification station is hanging for several seconds every 5 seconds.	Server
OCRProcessor.exe is crashed when processing the attached files	Server
It is not possible to select the particular CPU numbers in the properties of the Processing Station	Server
UserProperty is reset, if you move on to the next PageSlice	Server
Server does not apply the changes with the AD group without restarting the Server services.	Server
Scripting Demo. Need to add information.	Server
Scripting Demo. The formatting of the default scripts in the tabs General, Document Separation, Indexing and Output moved out.	Server
Outdated XmlResult.xsd and XmlTicket.xsd	Server
Export script may hang the jobs in the Publishing state.	Server
Attached file. The part of the text lost when exported to txt and xml formats	Synthesis
Processing error: Internal program error happens, when right-clicking on some words at the Verification Station	Verification
Japanese localization issues (IME utility issues, wrong fonts in text checking dialogs) at Verification Station	Verification

About the Product

About the Product Version

ABBYY Recognition Server 4 brings significantly improved recognition of Arabic text, new export options, processing of document libraries in both read-only and editable folders and other technology improvements. The new version comes with many revisions and upgrades in crucial areas such as server stability, performance, and auto-recovery. Other improvements include advanced logging, GUI changes and bug fixes. See below for details.

Installing the Product Version

Recognition Server 4 can be installed on the same computer as Recognition Server 3.5 or earlier versions.

Settings from an earlier version of ABBYY Recognition Server can be imported into ABBYY Recognition Server 4. For details, see the *Upgrade from the previous versions of ABBYY Recognition Server* chapter of the *System Administrator's Guide*.

Note: Recognition Server 4 includes changes to XML result files. If you are upgrading from version 3.5, this may require changes in the software used for integrating ABBYY Recognition Server with data storage systems. For details, see the *XML Result* section of the Help file.

License Usage

Recognition Server 4 does not work with most licenses generated for previous versions of Recognition Server (3.5 and earlier).

Some licenses that were generated for Recognition Server Arabic Edition can be used, but due to changes in license file parameters (the ISIS option has been added), we recommend generating new licenses for Recognition Server 4 Release 1 (for 3A), Recognition Server Release 1 and other maintenance releases.

History of Releases

Release 5 with Japanese Help Files Patch 1

Part #: **1135/23**, build # **4.0.6.4039**, OCR Technologies build # **13.0.28.139**, release date: **June 06, 2017**

- Correction of Japanese Administrator's Guide link in the Start menu.

Release 5 with Japanese Help Files

Part #: **1135/22**, build # **4.0.6.4037**, OCR Technologies build # **13.0.28.139**, release date: **May 30, 2017**

- Japanese localization of the help files of the Indexing Station and the Verification Station
- Office documents can now be processed using the web API
- Bug fixes

Release 5 Patch 1

Part #: **1135/21**, build # **4.0.6.126**, OCR Technologies build # **13.0.28.123**, release date: **February 02, 2017**

- Bug fixes of memory leak

Release 5

Part #: **1135/20**, build # **4.0.6.118**, OCR Technologies build # **13.0.28.117**, release date: **November 28, 2016**

- Improved e-mail processing
- Support for Microsoft SharePoint 2016
- Microsoft Failover Cluster support
- Bug fixes

Release 4 for Symantec

Part #: **1135/18**, build # **4.0.5.8891**, OCR Technologies build # **13.0.24.96**, release date: **September 01, 2016**

- Support for the **Symantec DLP Connector** custom license parameter

Release 4

Part #: 1135/14, build # 4.0.5.5022, OCR Technologies build # 13.0.24.96, release date: February 02, 2016

- Support of Microsoft SharePoint Online (Office 365)
- SharePoint library processing improvements
- Built-in component for conversion of digitally created documents

Release 3 with Japanese UI and Help

Part #: 1135/13, build # 4.0.4.1447, OCR Technologies build # 13.0.20.56, release date: October 9, 2015.

- Japanese localization of operator station UI and Help
- A bug fix for the ABBYY USA

Release 3 Patch 2 (for customer)

Part #: 1135/12, build # 4.0.4.1438, OCR Technologies build # 13.0.20.56, release date: September 22, 2015.

- Bug fix. Processing Stations can now be connected to the Server when the TCP/IP protocol is used for interactions between Recognition Server components. The “Access is denied” bug has been fixed.

Release 3 Patch 1 (for customer)

Part #: 1135/11, build # 4.0.4.1437, OCR Technologies build # 13.0.20.56, release date: September 07, 2015.

- Bug fix. Input files stored in Microsoft SharePoint libraries can now be overwritten with the output file when the two files have the same name and file extension. This prevents the duplication of documents

Release 3 with Japanese UI

Part #: 1135/10, build # 4.0.4.1434, OCR Technologies build # 13.0.20.56, release date: July 17, 2015.

- The UI of the Administration and Monitoring Console was translated into Japanese.

Release 3

Part #: 1135/9, build # 4.0.4.1425, OCR Technologies build # 13.0.20.54, release date: June 15, 2015.

- Conversion of documents in office formats
- Processing of entire SharePoint portals with child sites within one workflow
- Saving output files in input folders
- Adding original documents as attachments to PDF/A and PDF documents
- Improved export to ALTO XML
- Option to use SMTP servers for sending notifications to the Administrator

Release 2 Patch 2

Part #: 1135/8, build # 4.0.3.1180, OCR Technologies build # 13.0.15.138, release date: February 6, 2015.

- Improved work with multiple workflows (more than 16) and indexing of documents that contain many index fields.

Release 2 Patch 1

Part #: 1135/7, build # 4.0.3.1175, OCR Technologies build # 13.0.15.138, release date: January 16, 2015.

- Option to specify the fill color of empty space (“triangles”) on the edges of documents that have been automatically deskewed

Release 2

Part #: 1135/6, build # 4.0.3.1167, OCR Technologies build # 13.0.15.131, release date: November 14, 2014

- Improved MRC compression method (provides the best possible compression rates for PDF files)
- Option to use IFilter for processing PDF files in Microsoft SharePoint
- SharePoint library processing:
 - Crawling of the whole site (including multiple libraries and folders)
 - Options for setting up repeated crawling
- Export to specific column types in SharePoint (support of Date, Number, and other formats)
- Export to PDF/A-3

Release 1 Multilingual

Part #: 1135/5, build # 4.0.2.952, OCR Technologies build number 13.0.13.21, release date: 14/08/2014

- Localization of the UI and help to the following languages:
 - French
 - German
 - Italian
 - Spanish
 - Chinese
 - Portuguese (Brazil)
 - Czech
 - Hungarian
 - Polish
- Bug fix for ABBYY USA

Release 1

Part #: 1135/4, build # 4.0.2.943, OCR Technologies build number 13.0.13.15, release date: 19/05/2014

- Improved failure recovery
- Option to limit the number of processed pages
- Verification and Indexing Station improvements:
 - Selecting documents from a queue
 - Timeout settings
 - Saving changes on the stations
- Indexing Station improvements
 - Importing document types from an external source

Release 1 (specially for 3A)

Part #: 1135/3, build # 4.0.1.795, OCR Technologies build number 13.0.8.108, release date: 29/01/2014

- Improved server operation
 - Redundancy
 - Reports and statistics
- PDF file processing improvements
- Processing of documents in read-only folders
- Processing of documents in SharePoint libraries
- Latest technology version

Arabic Edition

Part #: 1135/2, build # 4.0.0.461, OCR Technologies build number 13.0.0.58, release date: 06/05/2013

- Improved recognition of Arabic texts
- Processing of documents in read-only folders
- Improved logging
- Bug fixes