

ABBYY FINEREADER ENGINE 11 FOR LINUX X64 FOR GLIBC 2.5 TECHNICAL RELEASE NOTES

CONTENTS

R1 GM	3
PART NUMBER, BUILD NUMBER	3
SOFTWARE AND HARDWARE REQUIREMENTS	3
NEW FEATURES AND IMPROVEMENTS	3
<i>Glibc 2.5 is supported</i>	3
<i>Java Samples</i>	3
FIXED BUGS	4
KNOWN ISSUES AND WORKAROUNDS	4
<i>Wrong list of supported OS and glibc versions in the Help file</i>	4
<i>BatchProcessor is not stable</i>	4
<i>Warnings during parallel processing</i>	4
<i>Activation scripts shows several unavailable modules</i>	4
LINUX VERSION LIMITATIONS	4
<i>Samples being installed under 'root' account are read only for others</i>	4
<i>IFontSet::EnablePdfStandardFonts is non-functional</i>	4
<i>OnPageProcessed comes once per a document</i>	5
<i>Some API is not implemented</i>	5
<i>Recognition of a PDF file with text in Vietnamese may fail</i>	5
<i>PDF/A validation report</i>	5
R2 GM	6
PART NUMBER, BUILD NUMBER	6
NEW FEATURES AND IMPROVEMENTS	6
<i>Java wrapper is included into the distribution</i>	6
<i>RPM is now used in distributives</i>	6
<i>IEngine::InjectTextLayer method was added</i>	6
<i>Support for IntelligentMail barcodes</i>	6
<i>Predefined languages OcrA and OcrB were added</i>	6
<i>Detection of vertical text for all languages</i>	6
<i>New IClassificationTrainer::AddPage method</i>	6
<i>The possibility to add custom features for classification</i>	7
<i>New predefined profile EngineeringDrawingsProcessing</i>	7

<i>Export with original layout to XLSX</i>	7
<i>New properties for IBarcodeParams object</i>	7
<i>Support for embedded files in PDF</i>	7
<i>New property IEngine::Version</i>	8
<i>Screenshot detection</i>	8
<i>Improved key value tables detection</i>	8
<i>Improved memory allocation</i>	8
<i>BatchProcessor acceleration on small images</i>	8
<i>Silent mode for runtime installation is now supported</i>	8
FIXED BUGS	8
KNOWN ISSUES AND WORKAROUNDS	10
<i>Linux version limitations</i>	10
<i>OnPageProcessed comes once per a document</i>	10
<i>Some API is not implemented</i>	10
<i>An error during the unloading of libFREngine.so on SLES 11 SP2 and SP3</i>	10
<i>Memory leak during processing of PDF files on Java</i>	11
<i>PDF/A validation report</i>	11

R1 GM –

PART NUMBER, BUILD NUMBER

Part#	1155/6
Build#	11.1.3.458651

SOFTWARE AND HARDWARE REQUIREMENTS

ABBYY FineReader Engine 11 for Linux is designed for glibc version 2.5 and above.

For the FineReader Engine dynamic library the standard libstdc++.so.6 and libgcc_s.so.1 libraries have to be used.

ABBYY FineReader Engine 11 for Linux has been tested on the following operating systems:

- Red Hat Enterprise Linux 5.9
- Oracle Enterprise Linux 5.9
- Ubuntu 12.04 LTS @ Amazon EC2

PC with Intel® Pentium® or compatible processor (1 GHz or higher) which supports SSE and SSE2 instruction sets.

Memory:

- for processing one-page documents — minimum 400 MB RAM, recommended 1 GB RAM
- for processing multi-page documents — minimum 1 GB RAM, recommended 1,5 GB RAM

Hard disk space: 800 MB for library installation and 100 MB for program operation plus additional 15Mb for every processing page of a multi-page document.

NEW FEATURES AND IMPROVEMENTS

GLIBC 2.5 IS SUPPORTED

This is the release with glibc 2.5 support.

JAVA SAMPLES

Hello sample for Java has been added.

This Hello sample demonstrates how to use ABBYY FineReader Engine in Java within JNI (Java Native Interface). It includes a C++ “HelloJNI” project with a sample JNI-wrapper implementing the most essential methods of FRDocument object.

The JNI-wrapper was ported from FineReader Engine 11 for Windows. It was tested with JDK 1.6, 1.7.

FIXED BUGS

There are no new fixes.

KNOWN ISSUES AND WORKAROUNDS

WRONG LIST OF SUPPORTED OS AND GLIBC VERSIONS IN THE HELP FILE

The list of supported operating systems and glibc versions was not updated.

BATCHPROCESSOR IS NOT STABLE

BatchProcessor could miss images in the processing queue in case there is not enough memory.

WARNINGS DURING PARALLEL PROCESSING

The following warnings could appear during parallel processing:

OMP: Warning #72: KMP_AFFINITY: affinity only supported for Intel(R) processors.

OMP: Warning #71: KMP_AFFINITY: affinity not supported, using "disabled".

These warnings occur on non-intel processor hosts and as a workaround it is recommended to set "export KMP_AFFINITY=disabled"

ACTIVATION SCRIPTS SHOWS SEVERAL UNAVAILABLE MODULES

There are some modules which are unsupported on Linux, but they are shown in the activation script.

LINUX VERSION LIMITATIONS

The following functionality of FineReader Engine 11 for Windows is not available in the Linux version:

- DjVu opening
- Scanning
- ICR/OMR
- Visual Components and other GUI elements
- WDP/WIC/BITMAP input formats and other Windows-specific functionality

SAMPLES BEING INSTALLED UNDER 'ROOT' ACCOUNT ARE READ ONLY FOR OTHERS

Samples are read-only for accounts with less than 'root' privileges if they are installed under 'root' account.

This will be fixed in the maintenance release.

IFONTSSET::ENABLEPDFSTANDARDFFONTS IS NON-FUNCTIONAL

It is not possible to use the standard PDF fonts during synthesis. That may lead to font embedding even though the text is typed using one of the standard PDF fonts.

This will be fixed in the maintenance release.

ONPAGEPROCESSED COMES ONCE PER A DOCUMENT

During opening, synthesis, exporting stages the callback OnPageProcessed arrives only once per a document rather than after each page.

This will be fixed in the maintenance release.

SOME API IS NOT IMPLEMENTED

The following API is not implemented in FRE 10 and FRE 11:

- IFootnoteSeries::IsNumberingWithSuperscript. Always returns “false”.
- IFootnoteSeries::PositionOnPage. Always returns “FPPT_SingleColumnSection”.
- IFootnoteSeries::PositionInDocument. Always returns “FPDT_PageEnd”.
- IFootnoteSeries::HasSeparator. Always returns “true”.
- ITextPicture::ColumnNumber. Always returns “0”.
- ICharParams::IsWordStart. Always returns “false”. It is true only for character parameters got through IWordRecognitionVariants interface.
- IIncut::TextWrapping. Always returns “TW_Undefined”.
- IRunningTitlesSeriesText::HasSeparator. Always returns “false”.

The implementation is not planned.

RECOGNITION OF A PDF FILE WITH TEXT IN VIETNAMESE MAY FAIL

Recognition fails with internal program error on certain PDF files with text layer in Vietnamese when “content reuse mode” is on.

This will be fixed in the maintenance release.

PDF/A VALIDATION REPORT

The following issues are known for PDF/A files produced by this release of FRE 11:

Adobe Acrobat 11.0.3 reports “Text cannot be mapped to Unicode” for 2% of images in CJK languages recognized and exported into PDF/A-1a or PDF/A-2a formats. On the other hand <http://www.pdf-tools.com/pdf/validate-pdf-a-online.aspx> on-line validator finds no issue in the same documents.

callas pdfaPilot, 3.1 (156) and 4 report “Image is not valid” for few images exported into PDF/A-2a (-2u) format. At the same time Adobe Acrobat 10.1.4, 11 report no issues with these files.

R2 GM –

PART NUMBER, BUILD NUMBER

Part#	1155/11
Build#	11.1.4.506920

NEW FEATURES AND IMPROVEMENTS

JAVA WRAPPER IS INCLUDED INTO THE DISTRIBUTION

From now on Java wrapper is implemented and entire SDK functionality is available through it. Now FRE can be easily integrated into Java environment.

RPM IS NOW USED IN DISTRIBUTIVES

RPM software package is available.

IENGINE::INJECTTEXTLAYER METHOD WAS ADDED

This is a new method which creates a copy of the input PDF file and adds the text layer which corresponds to the recognized text of the document.

SUPPORT FOR INTELLIGENTMAIL BARCODES

Starting with this release, ABBYY FineReader Engine recognizes IntelligentMail barcodes. Intelligent Mail is a height-modulated 65-bar barcode which is used on mail in the United States. It is also known as USPS 4-CB.

BT_ IntelligentMail was added to BarcodeTypeEnum enumeration constants.

PREDEFINED LANGUAGES OCRA AND OCRB WERE ADDED

The OCR A and OCR B modules now license not only corresponding text types but also the predefined languages which are provided for these text types. Apart from regular characters the alphabet for these languages also contains special symbols from OCR_A and OCR_B standards.

DETECTION OF VERTICAL TEXT FOR ALL LANGUAGES

IPageAnalysisParams::DetectVerticalEuropeanText property is a new property which allows you to detect vertical text in languages other than CJK (HelpDesk request 369221).

NEW ICLASSIFICATIONTRAINER::ADDPAGE METHOD

From now on FRE 11 supports `IClassificationTrainer::AddPage()` method.

This is a new method which allows you to add an already recognized page to the classification database.

THE POSSIBILITY TO ADD CUSTOM FEATURES FOR CLASSIFICATION

Now it is possible to add custom features that can enhance the accuracy of classification.

The following methods were implemented:

1. `IClassificationTrainer::AddFeaturesForPage()`

This is a new method which supports training the classification database with the help of user-defined additional features.

2. `IFRPage::ClassifyEx()`

This is a new method which supports classification with the help of user-defined additional features.

NEW PREDEFINED PROFILE ENGINEERINGDRAWINGSPROCESSING

This profile is specially developed for engineering drawings and maps recognition.

`EngineeringDrawingsProcessing` would be useful for text extraction disregarding its orientation and conversion of engineering drawings from PDF to searchable PDF.

EXPORT WITH ORIGINAL LAYOUT TO XLSX

New property `IXLExportParams::LayoutRetentionMode` that allows to export files to XLSX with original layout was added.

It is possible to choose the mode of retaining tables' formatting during export to XLSX.

`XLSXLayoutRetentionModeEnum` gives the ability to choose the mode that is better corresponding to the assigned task.

NEW PROPERTIES FOR IBARCODEPARAMS OBJECT

This release includes a new property, which allows you to detect more barcodes, but slows down the processing. `IBarcodeParams::EnableAdvancedExtractionMode` property enables the feature.

New `IBarcodeParams::MinRatioToTextHeight` property defines the minimal acceptable height of the barcode in relation to the average letters height. This setting can help to detect low barcodes.

SUPPORT FOR EMBEDDED FILES IN PDF

IPDFExportFeatures::WriteSourceAttachments property specifies if attachments from the original PDF file should be written to the output file. This property is ignored if the original file was not PDF.

New values PCM_Pdfa_3a and PCM_Pdfa_3u were added to enumeration constants PDFAComplianceModeEnum for PDF/A-3a and PDF/A-3u formats correspondingly .

NEW PROPERTY IENGINE::VERSION

IEngine::Version returns the build number of the ABBYY FineReader Engine version you are using.

SCREENSHOT DETECTION

From now on, our technologies are able to identify screenshots in documents. The document analyzer detects screenshots as image blocks if EnableTextExtractionMode property set to false (it is default value). Otherwise, text blocks can be detected on the screenshots (HelpDesk request 343728).

IMPROVED KEY VALUE TABLES DETECTION

Significant improvements were made for key value tables' detection in this release.

IMPROVED MEMORY ALLOCATION

The developers have successfully managed to significantly reduce memory fragmentation. Thus, the error OUT_OF_MEMORY was fixed in many cases.

BATCHPROCESSOR ACCELERATION ON SMALL IMAGES

From this release BatchProcessor works faster on the batches with small images.

SILENT MODE FOR RUNTIME INSTALLATION IS NOW SUPPORTED

In previous release it was only possible to launch the installation process in interactive mode. Now it possible to launch the installation process in automatic mode setting all the parameters via command line.

FIXED BUGS

This section contains a list of bugs reported by customers that have been fixed.

The following terminology is used for frequency in the list:

General bugs are frequently reproduced.

Rare bugs are seldom reproduced on certain conditions.

Specific bugs are reproduced in custom images in very specific scenarios.

Frequency	Description	Subsystem	HD #	Office
General	"Segmentation fault" occurs on customer's specific Virtual Machine	API	408410	UA
	The properties Title, Producer and DocumentInformationDictionary of DocumentContentInfo were not recorded during the opening of PDF file. Now the issue is fixed.	API	398171	EU
	The size of PDF file with Jpeg2000 compression was bigger than with Jpeg compression. Now the compression works correctly and PDF files with Jpeg2000 compression are smaller than PDF files with Jpeg compression.	API	380139	EU
	CommandLineInterface sample crashed after several launches of this sample.	Samples	388270	EU
	An error during export in CLI sample in case of export limitations in the license.	Samples	379964	US
	New information was added in the Help file that the detection of fonts depends on the installed fonts packages.	Help	392454	US
	There was an error during the registration of libraries in the system by sudo Inconfig command. The problem was in the linker scripts. Now the issue is fixed.	API	385322	EU
	The distributive didn't work on Linux host with 32 physical cores.	API	381584	EU
	IRTFExportParams::KeepPages property didn't work correctly. There was a disparity between the API and the Help file. Now this property is working as described in the Help file: the value of this property is used only if the PageSynthesisMode property is set to PSM_RTFFormatParagraphs or PSM_RTFFlaintext, otherwise it is ignored.	API	-	-
	An error occurred during the call GetEngineObjectEx with the property IsSharedCPUCoresMode = true.	API	-	-
Samples being installed under 'root' account were read only for other accounts. Now every user is able to modify the samples.	Samples	-	-	

KNOWN ISSUES AND WORKAROUNDS

LINUX VERSION LIMITATIONS

The following functionality of FineReader Engine 11 for Windows is not available in the Linux version:

- DjVu opening
- Scanning
- ICR/OMR
- Visual Components and other GUI elements
- WDP/WIC/BITMAP input formats and other Windows-specific functionality

ONPAGEPROCESSED COMES ONCE PER A DOCUMENT

During opening, synthesis, exporting stages the callback `OnPageProcessed` arrives only once per a document rather than after each page.

This will be fixed in the maintenance release.

SOME API IS NOT IMPLEMENTED

The following API is not implemented in FRE 10 and FRE 11:

- `IFootnoteSeries::IsNumberingWithSuperscript`. Always returns "false".
- `IFootnoteSeries::PositionOnPage`. Always returns "FPPT_SingleColumnSection".
- `IFootnoteSeries::PositionInDocument`. Always returns "FPDT_PageEnd".
- `IFootnoteSeries::HasSeparator`. Always returns "true".
- `ITextPicture::ColumnNumber`. Always returns "0".
- `ICharParams::IsWordStart`. Always returns "false". It is true only for character parameters got through `IWordRecognitionVariants` interface.
- `IIncut::TextWrapping`. Always returns "TW_Undefined".
- `IRunningTitlesSeriesText::HasSeparator`. Always returns "false".

The implementation is not planned.

AN ERROR DURING THE UNLOADING OF LIBFREENGINE.SO ON SLES 11 SP2 AND SP3

An error message is shown during the unloading of `libFREngine.so` on SLES 11 SP2 and SP3.

MEMORY LEAK DURING PROCESSING OF PDF FILES ON JAVA

Out of memory asserts and internal program errors periodically occur during processing of large amount of PDF files.

We recommend to use 4 GM RAM or more for parallel processing of large PDF documents.

PDF/A VALIDATION REPORT

The following issues are known for PDF/A files produced by this release of FRE 11:

1. Adobe Acrobat 11.0.5 (Preflight 11.0.4) detects an error for PDF/A with attachments (PCM_Pdfa_3a):

“Embedded file does not have AF entity”