

ABBYY FINEREADER ENGINE 11 FOR MAC

TECHNICAL RELEASE NOTES

CONTENTS

R1 GM	4
PART NUMBER, BUILD NUMBER	4
PRODUCT DESCRIPTION	4
WHAT IS NEW	4
<i>New Usage Scenarios</i>	4
<i>OCR Improvements</i>	7
<i>New Image Preprocessing Tools</i>	8
<i>Performance Improvements / Shortened Development Curve</i>	9
<i>Export Improvements</i>	10
<i>Useful and Clear Developer's Documentation</i>	13
<i>New Licensing</i>	13
<i>Other Improvements</i>	13
UPGRADE FROM PREVIOUS VERSION	14
<i>Installing on the Same Machine</i>	14
<i>Compatibility of Protection Keys</i>	14
<i>Using Source Code for Previous Version</i>	14
COMPATIBILITY ISSUES WITH VERSION 8	14
<i>Licensed 3rd-party software</i>	14
COMPONENTS DELIVERY	14
<i>FTP Delivery</i>	14
<i>DVD Box</i>	15
DISTRIBUTION COMPONENTS	15
<i>Documentation</i>	15
<i>Sample images</i>	15
<i>Code samples</i>	15
LICENSING MODEL AND PARAMETERS	16
PROTECTION KEY TYPES, ACTIVATION, DEACTIVATION, REGISTRATION	16
<i>Supported Protection Types</i>	16
LICENSE LIMITATIONS	16
<i>Page Counter</i>	16
SUPPORTED LANGUAGES, TYPES AND FORMATS	18
<i>Supported Recognition Languages</i>	18
<i>Supported barcode types</i>	19
<i>Supported text types and field marking types</i>	19
<i>Supported import and export formats</i>	20

GETENGINEOBJECT FUNCTION CHANGES.....	21
FULL NATIVE 64-BIT SUPPORT.....	21
CLASSIFICATION	21
<i>Classification modes</i>	21
<i>Classification confidence</i>	22
FLEXIFORMSDA AND FULLTEXTINDEXDA OPTIONS IN API AND LICENSING	22
CJK DOESN'T USE USER PATTERNS	22
SOFTWARE AND HARDWARE REQUIREMENTS.....	22
NEW FEATURES AND IMPROVEMENTS	23
<i>Support for Russian with Accents language</i>	23
<i>Support for Russian Old Spelling language</i>	23
<i>Support for FullAscii and Code32 1D barcodes</i>	23
<i>Re-formatted Distribution.csv file for easier runtime file list composing</i>	23
<i>New export formats are supported in CLI sample: ALTO, EBook</i>	24
<i>JBIG2 lossless compression is supported</i>	24
<i>Improvements in MRC visual quality</i>	24
<i>Method for checking if a page is empty</i>	25
<i>Aggressive table detection mode</i>	25
<i>Predefined processing profile for writing highly compressed image-only PDFs</i>	25
<i>XPS export format</i>	25
<i>Retaining document layout in textual export</i>	25
<i>Distribution is digitally signed</i>	26
FIXED BUGS	26
MAC OS VERSION LIMITATIONS.....	26
KNOWN ISSUES AND WORKAROUNDS	26
<i>The release supports only Open License protection</i>	26
<i>Java wrapper is not included into the distribution</i>	26
<i>No built-in parallel processing</i>	27
<i>IFontSet::EnablePdfStandardFonts is non-functional</i>	27
<i>OnPageProcessed comes once per a document</i>	27
<i>Some API is not implemented</i>	27
<i>PDF/A validation report</i>	27
R2 GM –	28
PART NUMBER, BUILD NUMBER.....	28
NEW FEATURES AND IMPROVEMENTS	28
<i>Support of activation for software licenses</i>	28
<i>IEngine::InjectTextLayer method was added</i>	28
<i>Support for IntelligentMail barcodes</i>	28
<i>Predefined languages OcrA and OcrB were added</i>	28
<i>Detection of vertical text for European languages</i>	28
<i>New IClassificationTrainer::AddPage method</i>	28
<i>The possibility to add custom features for classification</i>	29
<i>New predefined profile EngineeringDrawingsProcessing</i>	29
<i>Export with original layout to XLSX</i>	29
<i>New properties for IBarcodeParams object</i>	29

<i>New property IEngine::Version</i>	29
FIXED BUGS	30
KNOWN ISSUES AND WORKAROUNDS	30
<i>Mac OS version limitations</i>	30
<i>Java wrapper is not included into the distribution</i>	31
<i>No built-in parallel processing</i>	31
<i>OnPageProcessed comes once per a document</i>	31
<i>Some API is not implemented</i>	31
<i>PDF/A validation report</i>	32

R1 GM –

PART NUMBER, BUILD NUMBER

Part#	1161/7
Build#	11.1.3.447140

PRODUCT DESCRIPTION

ABBYY FineReader Engine 11 for Mac is a comprehensive software development kit (SDK) for integrating ABBYY's multilingual OCR, OBR, BCR, document classification, document imaging, document conversion and PDF conversion technologies into applications for Mac operating systems.

WHAT IS NEW

NEW USAGE SCENARIOS

Feature	Description	Benefits
<p>Automatic document classification</p> <p><i>The task of document classification is to assign a document to some category on base of its content.</i></p>	<p>ABBYY FineReader Engine 11 provides an API for automatic document classification that enables applications to categorize and sort batches of documents by predefined document classes.</p> <p>Classification in FineReader Engine 11 can be performed in two modes (classification profiles):</p> <ul style="list-style-type: none"> Maximum Speed. This mode is useful for documents that contain not much text, and the difference between classes is visible in the appearance of the documents. It uses the following classification criteria: <ul style="list-style-type: none"> Image pattern (black pixels location template) 	<p>Enables workflow automation</p> <p>Boost productivity and reduce costs by eliminating manual pre-sorting</p> <p>It's easy, no specific knowledge needed. No templates required. Can be trained easily by non-technical end-users.</p> <p>Easily adjustable. Whenever he needs anyone can train the engine to classify new types of documents.</p> <p>Universal. Fits for all types of documents.</p> <p>Easy to integrate. One</p>

	<ul style="list-style-type: none"> ○ OCRed text analysis: Title text ● Maximum accuracy. This mode is useful for documents that contain a lot of text, and the difference between classes can be determined only when text content is taken into account. Classification criteria are: <ul style="list-style-type: none"> ○ OCRed text analysis: Full-text <p>The “Maximum speed” mode provides classification speed from 3 to 10 times faster than “Maximum accuracy” mode.</p> <p>The classification results are:</p> <ul style="list-style-type: none"> ● Detected category of the document, ● Probability that a document belongs to a category. <p>Classification probability may be used to determine how to further process classified documents, for example, whether to re-classify some of documents manually, or to be able to route documents to the right department.</p> <p>Sample usage scenarios:</p> <ul style="list-style-type: none"> ● Archiving: Sorting documents by type for electronic archive creation. ● Mailrooms and Workflow Automation: According to document class detected some further actions can be initiated. ● Batch Processing: Document separation. ● BPO: Pre-sort documents for further processing. ● Banking/Insurance: Verification of document set completeness is applied to loan applications and insurance payouts. ● OEM: Smart MFP/scanner interfaces suggesting typical actions for each document class. 	<p>page of code is required for basic scenario implementation.</p>
<p>Business Card Recognition</p>	<p>Business card recognition technology now is integrated in FineReader Engine 11. The API provides a full set of features for business cards processing: from special preprocessing features to the API that provides access to extracted data.</p> <div data-bbox="435 1419 1073 1774" data-label="Image"> </div> <p>The following fields can be extracted:</p> <ul style="list-style-type: none"> ● Personal name 	<p>Fast and easy way to transfer and put your business contacts to work</p> <p>Superior accuracy of text and data processing and recognition</p> <p>Add more value to your application</p> <p>A must-have add-on for any CRM system</p>

- Company name
- Position in the company
- Company address
- Phone number
- Fax
- Mobile phone number
- E-mail
- Web site

Export to vCard format

- Recognized data can be saved in vCard format, which is often used to pass business card by e-mail or networks.


Business Card reading in 27 languages


- ChinesePRC, ChineseTaiwan, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Indonesian, Italian, Japanese, Korean, Norwegian, NorwegianBokmal, NorwegianNynorsk, Polish, PortugueseBrazilian, PortugueseStandard, Russian, Spanish, Swedish, Turkish, Ukrainian.

Auto-splitting of multiple business cards scanned as one page


- Multiple business cards can be detected on a page during processing.
- Multiple business cards scanned on one page can be split into several pages before processing.

OCR IMPROVEMENTS

Feature	Description	Benefits
<p>Arabic OCR</p> 	<p>FineReader Engine 11 includes Arabic OCR technology.</p>	<p>Expand your business in Arabic markets</p> <p>Put your finger on the pulse of Arabic world</p>
<p>CJK (Chinese, Japanese and Korean) OCR</p> 	<p>CJK OCR is now available for Mac users.</p>	<p>Enter to local Oriental markets to compete with domestic OCR vendors</p>
<p>New languages for OCR</p> 	<p>FineReader Engine 11 for Mac supports 27 new recognition languages.</p>	<p>Widest language support available in the industry – 202 languages for OCR</p>
<p>Maxicode barcode support</p> 	<p>Maxicode barcode is used for tracking and managing the shipment of packages (i.e. by UPS company).</p>	
<p>USPS 4CB barcode type*</p> <p><i>* still under development</i></p>	<p>USPS 4CB or IMB is a barcode used by USA post office.</p> 	
<p>Receipt recognition</p> <p>(new Receipt text)</p>	<p>This type of text is designed for recognizing sales receipts, invoices, etc. Unlike the other types, it is not concerned with the actual font of the text. Rather, it tells the recognizer that there may be text of low quality, mostly in monospaced or normal font. The typical receipt text</p>	

type)	can look like this: 	
New mode for low resolution scans	The special new recognition mode for low quality documents – old faxes, low resolution scans provides 20% higher accuracy for such documents than standard Normal mode	Now you can recognize the low quality documents that caused too many OCR errors before

NEW IMAGE PREPROCESSING TOOLS

Feature	Description	Benefits
Camera OCR <i>Preprocessing features for photographed documents</i>	<p>New version of FineReader Engine includes the following new and improved preprocessing features:</p> <ul style="list-style-type: none"> • Geometrical distortions correction (not only trapezium distortions as in previous version, but any type of geometrical distortions), • Auto-cropping • Background lightening • ISO noise removal 	<p>Better recognition results</p> <p>Allows to produce high quality searchable PDF with excellent appearance</p>
Auto-splitting of double-page spread	<p>Books are usually scanned as double-page spreads. This produces some difficulties for recognition (curved lines, scanning shadows). In the output document it is usually better to have the book page-by-page.</p> <p>New version of FineReader Engine 11 can perform page splitting automatically. This means higher effectiveness of image preprocessing (curved lines correction, scanning shadows removal).</p> 	<p>Better appearance of output document (page-by-page).</p>

New image preprocessing methods	<ul style="list-style-type: none"> • RemoveNoise Method. This method reduces the noise on the image. • EnhanceLocalContrast Method. This method increases the local contrast of the image. Such preprocessing may increase recognition quality of low contrast images. 	Better recognition results
New binarization	Previous OCR SDK version provided very high quality of binarization, but in some the most difficult cases it could commit errors and losses of information. New binarization technology ensures the whole text retention and prevents information losses even in difficult cases.	New binarization provides outstanding quality and guarantees keeping of valuable information resulted in better OCR accuracy
Color stamps and marks filtering	If there are some stamps or marks, made by pen, marker on document image they usually interfere text and decrease OCR results. That is why ABBYY FineReader Engine 10 includes special feature for color marks and stamps filtering and improving recognition accuracy.	An excellent feature for data capture systems allows preventing data losses from fields covered by stamps and color marks


PERFORMANCE IMPROVEMENTS / SHORTENED DEVELOPMENT CURVE

Feature	Description	Benefits
Predefined processing profiles for popular usage scenarios	<p>A lot of developers mentioned that ABBYY FineReader Engine API is the most powerful and full-functional among OCR SDKs. Now it becomes simpler with new profiles for the most popular recognition tasks. They are predefined with optimal parameters for easy start and guaranteed OCR quality without long-time manual tuning.</p> <p>Manual parameters setting is also available for any custom solutions</p>	You do not need to spend time studying a huge amount of SDK API's features and parameters. That is enough to choose the profile and ABBYY SDK sets all optimal values for maximal OCR efficiency.
Native 64-bit support	FineReader Engine 11 now provides native 64-bit support and can be used in 64-bit applications without any further development.	Eliminating difficulties with 64-bit applications development
Opening images	In FineReader Engine 8.0 for Mac the source images can be loaded	Results in increased

from memory* <i>* still under development</i>	into the Engine from files only. In the new version, it is possible to implement your own external image queue (custom image source), which will either return references to files on demand or provide the source images as a stream in memory.	flexibility, security and performance
Built-in multiprocessing recognition	A multi-page document will be processed on multiple CPU cores to increase processing speed.	Increased processing speed
Processing documents from import to export with Batch Processor* <i>* still under development</i>	BatchProcessor provides high efficiency usage of CPU cores of a computer thus reflecting the speed of processor. It performs all processing stages in parallel mode.	
Improved font management API	Font management is much more easier with FineReader Engine 11 – it provides a variety of predefined font filters which save developer from manual font specifying: <ul style="list-style-type: none"> • default set used by ABBYY FineReader • a set for European languages • a set for Chinese language • a set for Japanese language • a set for Korean language • a set for Arabic language • a set for Hebrew language • a set for Thai language • a set for Armenian language 	Extended access to the fonts used during document synthesis.
Mac OS Sandbox compliance	To prevent application security issues to compromise your system, it is possible to run them inside a Sandbox in OS X.	Increased security for Mac users.

EXPORT IMPROVEMENTS


Feature	Description	Benefits
Improved PDF Export		
Faster PDF Export	Export to PDF now is up to 2-3 times faster than in previous version of FineReader Engine 8 for Mac.	Converting images into searchable PDF is one


<p>PDF export profiles</p>	<p>ABBYY FineReader Engine 11 for Mac provides predefined profiles with optimal values for popular export variants:</p> <ul style="list-style-type: none"> • MaxQuality • Balanced • MinSize • MaxSpeed <p>With predefined PDF export profiles you automatically set optimal values for particular task.</p>	<p>of the most needed scenarios on the market.</p> <p>Better appearance and minimum size of the document.</p>
<p>Higher quality of PDF MRC</p>	<p>PDF MRC improvements include:</p> <ul style="list-style-type: none"> • higher background image compression, • contrast elements stay in foreground. <p>Higher background image compression reduces the size of output PDF MRC file for up to 50%.</p>	
<p>Stamps and written notes processing for PDF MRC *</p> <p><i>* still under development</i></p>	<p>Stamps and written notes can be placed to any part of documents and lay over sensitive information, which is needed to be extracted. Such marks are merged with the text during binarization and sensitive information can be lost.</p> <p>FineReader Engine 11 provides special preprocessing mode for such cases. The idea of this preprocessing mode is that an image is split into two layers: color and black-and-white. Black-and-white layer is used for recognition, while the color layer is passed to export without modifications and it will stay in foreground of resulting PDF MRC file instead of appearing in background as blurred elements.</p> <p>As a result, output PDF has high-level quality and compression. Text and separators are compressed with suitable black-and-white codec, while color layer - with color codec.</p> 	

<p>An ability to create linear or non-linear PDF files</p>	<p>Linear PDF files have internal data arranged in a page order. A page of a linear PDF file can be read in a web browser plug-in without waiting for the whole file to be downloaded. Non-linear PDF have the data necessary to assemble a document page scattered through the whole file.</p> <p>New PDF export parameters include new option which specifies whether a linear PDF file should be created.</p>	
---	--	--

Other Export Improvements

<p>Docx and Xlsx Export Improvements*</p> <p><i>* based on ABBYY internal test results</i></p>	<p>Docx improvements:</p> <ul style="list-style-type: none"> • Better pages number detection • Better font detection: Font type , Bold, Italic, Header • Better text wrapping • Better page orientation detection <p>Xlsx improvements:</p> <ul style="list-style-type: none"> • Better cells detection • Better columns' width detection 	
---	---	--

<p>Excel export improvements*</p> <p><i>* still under development</i></p>	<p>FineReader Engine 11 allows retaining formatting of all data in the tables exported to Excel, including numbers:</p> <ul style="list-style-type: none"> • Bold font style • Font colors 	
--	---	--

<p>Export to XPS*</p>  <p><i>* still under development</i></p>	<p>XPS (XML Paper Specification) format is based on XML. As PDF format, it provides device-independent document appearance. XPS also comes handy when one doesn't have a printer installed and the XPS virtual printer allows to save the document in "ready-to-print" original format for later printing.</p>	
--	--	--

Export to memory* <i>* still under development</i>	FineReader Engine 11 will be able to save recognized documents not only on disk, but into a file stream.	Increased security for your confidential data
Extended ABBY XML	Now there is an ability to save paragraph style and roles into output XML. This can be useful to identify the role of a paragraph, e.g. to detect running titles and footnotes.	
Other	<ul style="list-style-type: none"> • Recreation of the logical structure of a document is an option during export to RTF, DOCX, and HTML formats. • An ability to save information of paragraph styles and roles in XML file. • New color settings for embedded pictures in RTF, DOCX, PPTX, HTML, EPUB, and FB2 formats. 	

USEFUL AND CLEAR DEVELOPER'S DOCUMENTATION

Improved Developers Guide (Help)	Improved Help distinguishes by updated structure and appearance together with new content including general product description, API specification, usage samples and best practices.	You will fast and easily find all necessary information and will enjoy working with ABBYY FineReader Engine 11.
---	---	---

NEW LICENSING

Distributed licensing	a license is set up in a local network for easy management	
Redundancy	ABBYY licensing service could be doubled for higher fault-tolerance	

OTHER IMPROVEMENTS

Feature	Description	Benefits
Opening PDF files from memory	FineReader Engine 10 can open image files in different formats from memory, but not PDF. PDF files must be saved to a disk before they can be processed with Engine. In FineReader Engine 11 this restriction	<ul style="list-style-type: none"> • Increased processing speed

	is removed.	
--	-------------	--

UPGRADE FROM PREVIOUS VERSION

INSTALLING ON THE SAME MACHINE

ABBYY FineReader Engine 11 works with any previous ABBYY FineReader Engine major version installed on the same PC if products' installation folders are different.

COMPATIBILITY OF PROTECTION KEYS

ABBYY FineReader Engine 11 requires valid serial number for functioning.

USING SOURCE CODE FOR PREVIOUS VERSION

ABBYY FineReader Engine 11 has certain incompatibilities with API implemented in the previous versions described in "ABBYY FineReader Engine 11 and 8.0 Compatibility" section of the product Help file. Every existing customer willing to upgrade his copy of ABBYY FineReader Engine should read the article first.

COMPATIBILITY ISSUES WITH VERSION 8

LICENSED 3RD-PARTY SOFTWARE

This version uses several licensed 3rd-party libraries. They enable the product with useful functionality and require us to add certain acknowledgements and items in the product documentation and/or LA.

The list of newly licensed technologies is below.

JPEG 2000 KAKADU LIBRARY

11th version uses JPEG 2000 Kakadu library for saving image files in JPEG 2000 format, or to export to PDF format with embedded JPEG 2000 pictures. That obliges us to specify certain copyrights in the product documentation:

- Working with JPEG2000 image format:
Portions of this software are copyright ©2011 University of New South Wales All rights reserved.

COMPONENTS DELIVERY

FTP DELIVERY

FTP delivery is mostly used for trial versions but can be also used for usual sales. It includes:

1. FTP address of the Distribution Pack.
2. Serial Number.

DVD BOX

It includes:

1. Common DVD Box with DVD Box Cover
2. CD/DVD with the Distribution Pack copy and the CD/DVD Label.
3. Serial Number.

DISTRIBUTION COMPONENTS

DOCUMENTATION

Material	Language	File name	Description
User's Guide	English	./Help/FREngine11UserGuide.pdf	Printing version of the Help File.
HTML Help	English	./Help/FREngine/index.htm	A full and detailed description of the product functionality. It also includes chapters on License Manager.
Distribution List	English	./Help/FREngine11_Distribution.csv	A list of files to distribute with description of responsibility and necessity to distribute.

SAMPLE IMAGES

There are prepared sample images for demonstration of basic scenarios and advanced technologies.

The distribution DVD contains all images in the following folder:

- DVD .\Samples

Sample images	Description
For demonstration of basic scenarios	Multi-page sample image in English. Available in the root of the SampleImages folder.
For demonstration of business card recognition (NEW)	Images of single and multiple business cards scanned on one page. Available in the root of the SampleImages folder.

CODE SAMPLES

The distribution contains samples described below in the following folder:

- DVD .\Samples

SAMPLES FOR DEVELOPERS

Name	Available in	Description
TiffToTf	<ul style="list-style-type: none">• Objective C	Performs document conversion with just a few lines of code. This sample will help you to start development using ABBYY SDK.

SAMPLES FOR TECHNOLOGY ADVANTAGES DEMONSTRATION

Name	Available in	Description
BusinessCardRecognition (New)	<ul style="list-style-type: none">Objective C	The sample shows how FREngine can extract data from business cards and illustrates how several business cards scanned on one page can be split.

LICENSING MODEL AND PARAMETERS

Licensing model and parameters are the same as for FineReader Engine 11 for Windows with the following exceptions:

- Hardware licenses are not supported
- The following Functional License Limitations are not available:
 - Data Capture (ICR/OMR): ICR, Cyrillic ICR, OMR
 - Visual Components

See details in the similar document for Windows version.

PROTECTION KEY TYPES, ACTIVATION, DEACTIVATION, REGISTRATION

SUPPORTED PROTECTION TYPES

1. Software protection keys:
 - a. Open (no activation).

LICENSE LIMITATIONS

PAGE COUNTER

Productivity Licensing means counting of processed pages or characters.

Engine treats a page as processed upon a call of the following methods:

- IFRDocument::
 - Analyze
 - AnalyzeAndRecognize
 - AnalyzeAndRecognizePages
 - AnalyzePages
 - Process
 - Recognize
 - RecognizePages
- IFRPage::
 - Analyze
 - AnalyzeAndRecognize
 - AnalyzeRegion
 - AnalyzeTable
 - Recognize

- RecognizeBlocks
- ExtractBarcodes
- IDocumentAnalyzer::
 - AnalyzePage
 - AnalyzePages
 - AnalyzeRegion
 - AnalyzeTable
 - RecognizePage
 - RecognizePages
 - RecognizeBlocks
 - AnalyzeAndRecognizePage
 - AnalyzeAndRecognizePages
 - RecognizeImageDocumentAsPlainText
 - ExtractBarcodes
- IEngine::
 - RecognizeImageFile
 - RecognizeImageAsPlainText
 - RecognizeImageDocumentAsPlainText
 - AnalyzePage
 - AnalyzePages
 - RecognizePage
 - RecognizePages
 - AnalyzeAndRecognizePage
 - AnalyzeAndRecognizePages

A page counter is increased by one only once for the same ImageDocument object regardless to how many times one uses analysis, recognition, or exporting methods with that object.

In case of character counter, the following methods have no effect:

- IDocumentAnalyzer::
 - AnalyzePage
 - AnalyzePages
 - AnalyzeRegion
 - AnalyzeTable
- IEngine::
 - AnalyzePage
 - AnalyzePages
- IFRDocument::
 - Analyze
 - AnalyzePages
- IFRPage::
 - Analyze
 - AnalyzeRegion
 - AnalyzeTable

SUPPORTED LANGUAGES, TYPES AND FORMATS

SUPPORTED RECOGNITION LANGUAGES

OCR LANGUAGES

	With dictionaries	Without dictionaries	Overall count
Common (default) languages	40	145 , including <ul style="list-style-type: none"> • 4 artificial languages <ul style="list-style-type: none"> ○ Esperanto ○ Ido ○ Interlingua ○ Occidental • 2 special languages (included by default if corresponding text type is chosen) <ul style="list-style-type: none"> ○ CMC7 ○ E13B • 6 programming languages • Chemistry • Digits 	185 , included in Runtime Professional
Additional languages	12 , including: <ul style="list-style-type: none"> • Arabic • Japanese • Hebrew • Korean • Korean Hangul • Thai • Vietnamese • FR XIX <ul style="list-style-type: none"> ○ Old English ○ Old French ○ Old German ○ Old Italian ○ Old Spanish 	5 additional languages <ul style="list-style-type: none"> ○ Chinese Simplified (PRC) ○ Chinese Traditional (Taiwan) ○ Yiddish (under Hebrew Add-On) ○ FR XIX <ul style="list-style-type: none"> ○ Old Slavonic ○ Latvian Gothic 	17 , included in Add-Ons
Total	52	150	202

CLASSIFICATION LANGUAGES

FineReader Engine supports all languages for classification.

BCR LANGUAGES

27 languages including:

- **4** hieroglyphic languages (CJK) - Chinese Traditional (Taiwan), Chinese Simplified (PRC), Japanese, Korean

SUPPORTED BARCODE TYPES

1D Barcodes			2D Barcodes	Overall count
15: <ul style="list-style-type: none"> ○ Codabar ○ Code 128 ○ Code 39 ○ Code 93 ○ EAN 8 ○ EAN 13 ○ IATA 2 of 5 ○ Industrial 2 of 5 ○ Interleaved 2 of 5 ○ Matrix 2 of 5 ○ Patch ○ PostNet ○ UCC-128 ○ UPC-A ○ UPC-E 	Including 4 with checksum: <ul style="list-style-type: none"> ○ Code 39 ○ Interleaved 2 of 5 ○ Codabar ○ Matrix 2 of 5 	Including 4 with supplemental: <ul style="list-style-type: none"> ○ EAN 8 ○ EAN 13 ○ UPC-A ○ UPC-E 	5: <ul style="list-style-type: none"> ○ PDF417 ○ Aztec ○ DataMatrix ○ QR Code ○ MaxiCode 	20

SUPPORTED TEXT TYPES AND FIELD MARKING TYPES

OCR TEXT TYPES

1. Normal
2. Fax
3. Typewriter
4. Matrix
5. OCR_A
6. OCR_B
7. MICR_E13B
8. MICR_CMC7
9. Fraktur/Gothic (available only under FineReader XIX add-on)
10. Receipt

FIELD MARKING TYPES

1. Simple Text
2. Underlined Text
3. Text in Frame
4. Grey Boxes
5. Char Box Series
6. Simple Comb
7. Comb in Frame
8. Partitioned Frame

SUPPORTED IMPORT AND EXPORT FORMATS

SUPPORTED IMPORT FORMATS

Format	Open	Save
BMP: uncompressed black and white 4- and 8-bit — uncompressed Palette 16-bit — uncompressed, uncompressed Mask 24-bit — uncompressed 32-bit — uncompressed, uncompressed Mask	+	+
BMP: 4- and 8-bit — RLE compressed Palette	+	
DCX: black and white 2-, 4- and 8-bit palette 24-bit color	+	+
GIF: black and white — LZW-compressed 2-, 3-, 4-, 5-, 6-, 7-, 8-bit palette — LZW-compressed	+	
JBIG2: black and white	+	+
JPEG: gray, color	+	+
JPEG 2000: gray — Part 1 color — Part 1	+	+
PCX: black and white 2-, 4- and 8-bit palette 24-bit color	+	+
PDF (version 1.7 or earlier)	+	+
PNG: black and white, gray, color	+	+
TIFF: black and white — uncompressed, CCITT3, CCITT4, Packbits, ZIP, LZW gray — uncompressed, Packbits, JPEG, ZIP, LZW 24-bit color — uncompressed, JPEG, ZIP, LZW 1-, 4-, 8-bit palette — uncompressed, Packbits, ZIP, LZW (including multi-page TIFF)	+	+
TIFF: black and white — CCITT3FAX	+	

SUPPORTED EXPORT FORMATS

1. RTF
2. Microsoft Office file formats:
 - a. DOCX
 - b. XLS/XLSX
 - c. PPTX

3. PDF file formats
 - a. PDF
 - b. PDF/A (1b, 1a, and 2a and 2u)
 - c. MRC (Mixed Raster Content) for both PDF and PDF/A
4. HTML
5. TXT/CSV
6. ABBYY XML
7. EPUB, ALTO, FB2
8. ODT
9. vCard — for export of business cards only

GETENGINEOBJECT FUNCTION CHANGES

The functions and methods, which load the Engine object, have their syntax changed:

The **GetEngineObject** function and **IEngineLoader::GetEngineObject** method do no longer have the parameters for **Open License**. To use Open License, one should use the **GetEngineObjectEx** function or **IEngineLoader::GetEngineObjectEx** method.

The **GetEngineObjectEx** function and **IEngineLoader::GetEngineObjectEx** method have one more additional parameter, which specifies whether CPU cores should be used in shared mode.

FULL NATIVE 64-BIT SUPPORT

FineReader Engine 11 provides 64-bit versions of libraries, including libraries for classification and BCR. Details on the distribution package can be found in the Help file.

CLASSIFICATION

Classification is a new feature in FineReader Engine 11. You can find the detailed description of the feature in marketing materials of the product. Below are several implementation details.

CLASSIFICATION MODES

Classification in FineReader Engine 11 can be performed in two modes:

- **Fast.** This mode is useful for documents that contain not much text, and the difference between classes is visible in the appearance of the documents. This mode uses image pattern (black pixels location template) and recognized titles of a document for classification.
- **Quality.** This mode is useful for documents that contain a lot of text, and the difference between classes can be determined only when text content is taken into account. This mode uses full text OCR for classification.

The mode is specified in **IClassificationParams::ClassificationMode** property.

CLASSIFICATION CONFIDENCE

The results of classification provide information both on the detected category of the document and the confidence, probability that a document belongs to this category (**IClassificationClass::Confidence**). There is also a flag (**IClassificationClasses::IsSuspicious**) indicating whether classification of a document was uncertain, e.g. if classification detected two classes with equal confidence for one document. One may use these values to determine the way of further processing for the classified documents, for example, re-classify some of the documents manually, if necessary.

FLEXIFORMSDA AND FULLTEXTINDEXDA OPTIONS IN API AND LICENSING

Starting from FREngine 11, **DA for Invoices** and **DA for Full-text Indexing** add-ons are free of charge and removed from licensing scheme. This is because of the following reasons:

- “TextExtraction” and “DocumentArchiving” predefined profiles require these add-ons. That is because the add-on modules were initially designed for those usage scenarios and therefore were included into corresponding profiles. Without add-ons Engine performs not optimally in those scenarios.
- Indexing and text extraction years ago were ‘additional’ usage scenarios for FineReader Engine while ‘document conversion’ was main one. It seems that all three scenarios are basic now for Engine.

Changes in licensing required additional changes in API:

- We declare the **FlexiFormsDA** and **FullTextIndexDA** properties as obsolete. These properties will be removed in the next version of the product.
- The old FlexiFormsDA property is replaced with **IPageAnalysisParams::EnableTextExtractionMode** and **IObjectsExtractionParams::EnableAggressiveTextExtraction** properties.
- The old FullTextIndexDA property is replaced with **IObjectsExtractionParams::DetectTextOnPictures** property.

CJK DOESN'T USE USER PATTERNS

CJK recognizer does not use cache so it is impossible to teach it with user patterns. The same situation is in ABBYY FineReader Engine 8.

SOFTWARE AND HARDWARE REQUIREMENTS

Operating System: Mac OS X (10.6.x, 10.7.x, 10.8.x).

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: 1 GB for library installation and 100 MB for program operation plus additional 15Mb for every processing page of a multi-page document.

5. `RequiredByModule` — the values in this column are equal to `Stage.Part.Details`, and there is no need to filter this column if the first three have been specified correctly. But it can be used to check which modules have been included.
6. `RequiredByInterfaceLanguage` — the interface language for which the file is necessary. The files marked "Any" are necessary independent of language settings.
7. `RequiredByRecognitionLanguage` — the recognition language for working with which the file is necessary. The files marked "Any" are necessary independent of recognition language.
8. `Optional` — specifies if the file is necessary for the module functionality. If the value is `No`, this file must be included in your distribution kit. The value can be set to `Yes` in the following cases:
 - 8.1. the file is language-specific. Include it if you need this language (consult columns 6 and 7).
 - 8.2. the functionality for which this file is responsible is not always necessary. For example, it can be used for opening images in a specific format. Consult ABBYY FineReader Engine Distribution Kit for further information about this file which will help you decide if you need it.

Finally, you receive the list of files. In the last three columns you will find the information about these files' location and size:

9. `Path` — file path in the root installation folder.
10. `FileName` — file name.
11. `Size` — file size in bytes.

NEW EXPORT FORMATS ARE SUPPORTED IN CLI SAMPLE: ALTO, EBOOK

CLI sample was updated to support additional export formats: ALTO and EBook.

JBIG2 LOSSLESS COMPRESSION IS SUPPORTED

This release includes new compression format for black and white images – JBIG 2 lossless.

It is useful if somebody works with low quality b/w images where lossy format may lead to similar characters substitution, e.g. "6" to "8" and vice versa.

This format is available for choosing in `PDFPictureCompressionParams::BwPictureFormats`.

IMPROVEMENTS IN MRC VISUAL QUALITY

It is possible to use black and white image prepared during binarization process as a source for MRC mask. In this case all contrast graphical elements will be saved into PDF MRC file with the highest quality.

This technique allows saving handwritten data, stamp details, signatures, etc. in a foreground layer with high quality (low compression) and keep a document look as close to original as possible. Previously enumerated objects were placed into background layer with high compression.

The following PDF export parameter enables the feature: `PDFMRCParams::UseBwImageAsTextMask`.

METHOD FOR CHECKING IF A PAGE IS EMPTY

FRPage::IsEmpty() method checks if the page is empty. It uses the same parameters as analysis methods to find out if the page contains any relevant objects, for example text, tables, or pictures.

If there is a high probability that some of your pages contain only barcodes, and they are relevant for your procedure, set the boolean NeedCheckBarcodes parameter to TRUE to help the method detect barcode-only pages.

This method is useful in a batch scanning scenario when there is a need to separate scanned images flow into documents. In this scenario a batch contains separating paper sheets which are blank or with certain barcodes printed on them.

AGGRESSIVE TABLE DETECTION MODE

PageAnalysisParams::AggressiveTableDetection property manages the table detection mode. If you set it to TRUE, FineReader Engine tries to find as many tables as possible on the page. This setting is recommended only for the documents which contain a lot of tables.

This property is FALSE by default.

PREDEFINED PROCESSING PROFILE FOR WRITING HIGHLY COMPRESSED IMAGE-ONLY PDFS

The Engine got new processing profile "*HighCompressedImageOnlyPdf*" for creating highly compressed image-only PDF files.

It is suitable for creating high-compressed PDF files which contain entire documents saved as pictures. The following settings are used:

- Document recognition and synthesis of the logical structure of a document are not performed.
- Skew correction is not performed.
- PDF export is optimized for the minimum size of the resulting file.
- The entire document is saved as a picture (PEM_ImageOnly mode).

XPS EXPORT FORMAT

Starting from this release FRE supports exporting to XPS format.

This format is developed by Microsoft in contrast to PDF format ([see Wiki](#)).

This is optional export format and to enable it one should enable this format in a license.

RETAINING DOCUMENT LAYOUT IN TEXTUAL EXPORT

TextExportParams::RetainLayout property turns on the export mode in which the original layout is simulated by inserting spaces. When displayed with a monospace font, the text and table columns will be level. The distance between blocks will be approximately the same.

This mode is not intended for right-to-left or vertical texts.

When this property is set to TRUE, the InsertEmptyLineBetweenParagraphs property is ignored, and the presence of empty lines is determined by the size of empty space between paragraphs in the original. The ExportParagraphsAsOneLine property is also ignored, and the line breaks are always kept.

This property is FALSE by default.

DISTRIBUTION IS DIGITALLY SIGNED

The distribution is signed with “Developer ID Application” certificate. This type of certificate is used to sign a Mac app before distributing it outside the Mac App Store.

FIXED BUGS

There are no bugs that have been reported by Beta testers and fixed in this release.

MAC OS VERSION LIMITATIONS

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

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

KNOWN ISSUES AND WORKAROUNDS

THE RELEASE SUPPORTS ONLY OPEN LICENSE PROTECTION

R1 GM works with Open Licenses only.

Software licenses are planned for R2 release.

JAVA WRAPPER IS NOT INCLUDED INTO THE DISTRIBUTION

Though we have partly functional Java wrapper for the Engine it still misses some important parts.

In some cases current version of the wrapper is enough, so please consult with HQ product analyst in case of urgent need.

We are planning to include fully-functional Java wrapper into R2 release.

NO BUILT-IN PARALLEL PROCESSING

The release does not include built-in multi-page document processing on multiple CPU cores.

The feature is planned to be included into R2 release.

IFONTSET::ENABLEPDFSTANDARD FONTS 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.

PDF/A VALIDATION REPORT

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

1. Adobe Acrobat 11.0.3 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.
2. 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.

PART NUMBER, BUILD NUMBER

Part#	1161/10
Build#	11.1.4.503210

NEW FEATURES AND IMPROVEMENTS

SUPPORT OF ACTIVATION FOR SOFTWARE LICENSES

Previous release worked only with Open Licenses that were originally intended for runtime installations. Now it is possible to activate software licenses via Internet or by email. From now on software licenses with activation support will be used for Developer and Developer Trial Licenses as this type of protection is more appropriate for these licenses.

Open Licenses will be used only for runtime installations as it is in Windows and Linux versions.

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 can recognize 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.

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 EUROPEAN 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.

NEW PROPERTY IENGINE::VERSION

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

FIXED BUGS

Below you can see a list of higher priority bugs that have been found by our testers and fixed in this release.

The following terminology is used for frequency in the list:

General bugs: frequently reproduced.

Rare bugs: seldom reproduced on certain conditions.

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

Frequency	Description	Subsystem	HD #	Office
General	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_RTFFormatPlainText, otherwise it is ignored.	API	-	HQ
Rare	A license wasn't displayed in the License Manager after successful activation.	Protection	-	HQ
Rare	An error in the License Manager. Steps to reproduce: 1. Click to "Activate license" 2. Choose an option "I already have activation file" 3. Input an activation file name An error message appears «License file was created for another computer".	Protection	-	HQ
Rare	A test failure on an English batch during preprocess stage.	API	-	HQ

KNOWN ISSUES AND WORKAROUNDS

MAC OS VERSION LIMITATIONS

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

- DjVu opening
- Scanning

- ICR/OMR
- Visual Components and other GUI elements
- WDP/WIC/BITMAP input formats and other Windows-specific functionality
- PDF text layer reusing
- Attachments extraction from PDF
- Bookmarks extraction from PDF
- Metadata extraction from PDF (the information about author, keywords, subject and title)

JAVA WRAPPER IS NOT INCLUDED INTO THE DISTRIBUTION

Though we have partly functional Java wrapper for the Engine it still misses some important parts.

In some cases current version of the wrapper is enough, so please consult with HQ product analyst in case of urgent need.

NO BUILT-IN PARALLEL PROCESSING

The release does not include built-in multi-page document processing on multiple CPU cores.

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.

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.

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 reports “Text cannot be mapped to Unicode” for 2% of images in CJK languages recognized and exported into PDF/A-1a, PDF/A-2a and PDF/A-3a 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.
2. Adobe Acrobat 11.0.5 reports “Implementation limit: Width of MediaBox below 3 units” for a specific document in Arabic recognized and exported into PDF/A-2a and PDF/A-3a formats. This issue is reproduced only on one document and is related to the problems of the input file.

PART NUMBER, BUILD NUMBER

Part#	1161/21
Build#	11.1.14.686241

UPGRADE FROM THE PREVIOUS VERSIONS AND RELEASES

BINARY INCOMPATIBILITY

It is necessary to recompile host application regardless a version of Engine previously used.

API CHANGES

IRECOGNIZERPARAMS.DETECTLANGUAGE PROPERTY IS MARKED AS DEPRECATED

New property `IRecognizerParams::LanguageDetectionMode` is now used instead of `IRecognizerParams::DetectLanguage`. `IRecognizerParams::DetectLanguage` is marked as deprecated.

For purposes of compatibility, new property value is changed whenever you change the `DetectLanguage` property: if you change it to `TRUE`, the property is set to `TSPV_Yes`, if you change it to `FALSE`, this property is changed to `TSPV_No`.

SOME PROPERTIES RELATED TO DOCUMENT STRUCTURE ARE MARKED AS DEPRECATED

The following properties are marked as deprecated:

- `IPageStructure::IsPageOdd`
- `IFootnoteSeries::HasSeparator`
- `IFootnoteSeries::PositionInDocument`
- `IFootnoteSeries::PositionOnPage`
- `IFootnoteSeries::IsNumberingWithSuperscript`
- `ISynthesisParamsForPage::InsertEmptyParagraphsForBigInterlines`

These methods are not implemented and the implementation is not planned.

HFM_FORMAT32 FORMAT IS MARKED AS OBSOLETE

From this release this format is obsolete. Using this constant is equivalent to `HFM_Format40` with the `IHTMLExportParams::UseCss` property set to `FALSE`.

ICHARPARAMS::ISWORDSTART IS DEPRECATED

ICharParams::IsWordStart property is deprecated and will be removed in the next major version. IsWordLeftmost property must be used instead.

CHANGES IN BEHAVIOR

CHANGES IN LANGUAGE AUTO DETECTION

In previous release property IRecognizerParams.DetectLanguage was set by default to FALSE.

In this release new property IRecognizerParams::LanguageDetectionMode is set to TSPV_Auto.

You can find detailed information about new IRecognizerParams::LanguageDetectionMode property in the section *New advanced language detection mode*.

PMINLIU AND MINGLIU FONTS ARE ALWAYS EMBEDDED DURING EXPORT TO PDF FILE

You can find detailed information in Correct display of PDF files with PMingLIU/MingLIU fonts in viewers section.

SUPPORT OF CORRUPTED TIFF FILES OPENING

Starting with this release, FineReader Engine can open corrupted TIFF files. It can be done only if one last line is corrupted, otherwise an error occurs and the corrupted file is not opened.

If the TIFF image which has been opened is corrupted, a warning message containing the number of the damaged page in the image file will be shown.

HIGHER OBR ACCURACY IN TRADE OF SPEED

'BarcodeRecognition_Accuracy' processing profile became more accurate in exchange of slowdown. For details see "BarcodeRecognition_Accuracy' became more accurate' below.

NEW FEATURES AND IMPROVEMENTS

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.

IMPROVED KEY VALUE TABLES DETECTION

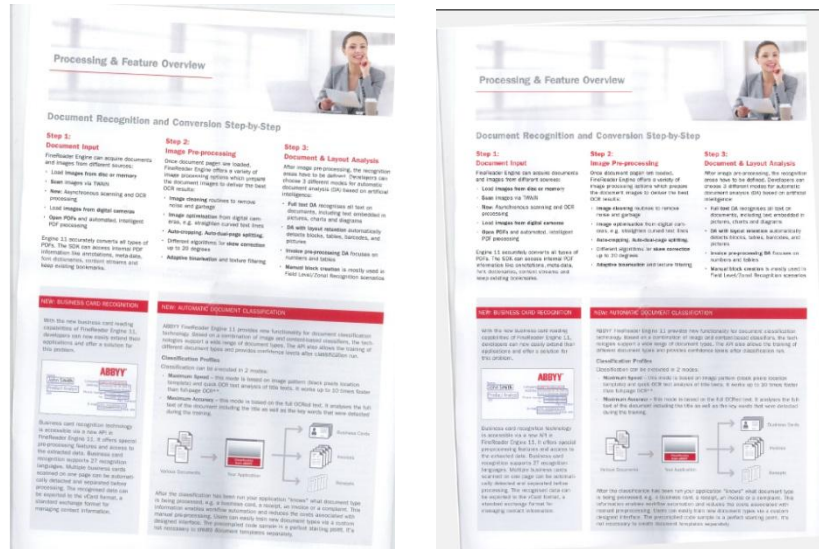
Significant improvements were made for key value tables' detection in this release.

POSSIBILITY TO ADD ATTACHMENTS TO PDF

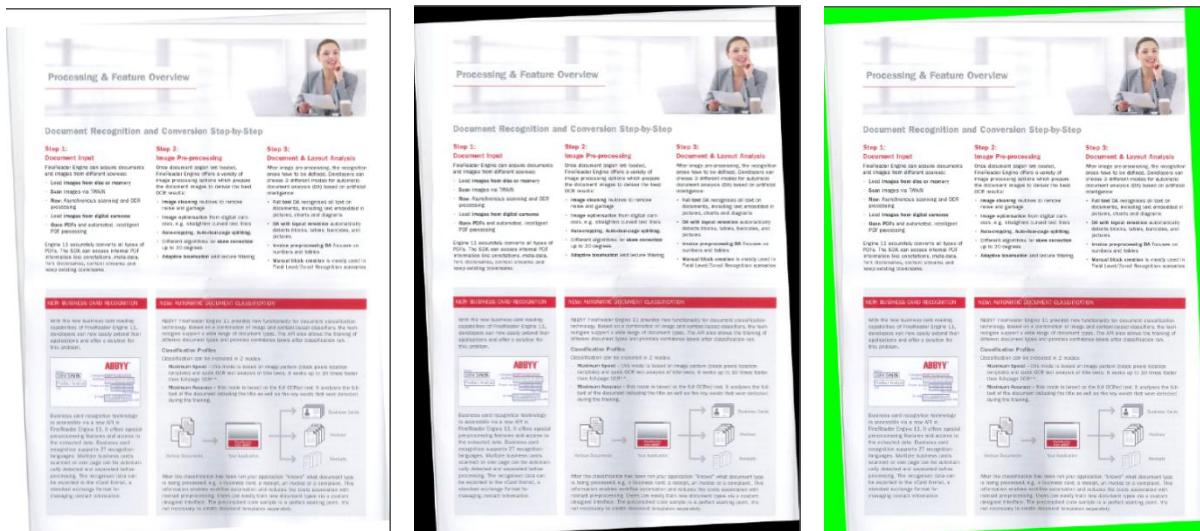
All the attachments are contained in **FRDocument::PDFAttachments**. You can add files to be attached to the output PDF file during export.

NEW PROPERTY IPREPREIMAGEMODE::BACKGROUNDFILLINGCOLOR TO FILL AREAS ADDED AFTER SKEW CORRECTION

After skew correction supplementary areas are added to the edges of document. The color of these areas is set automatically by Fine Reader Engine:



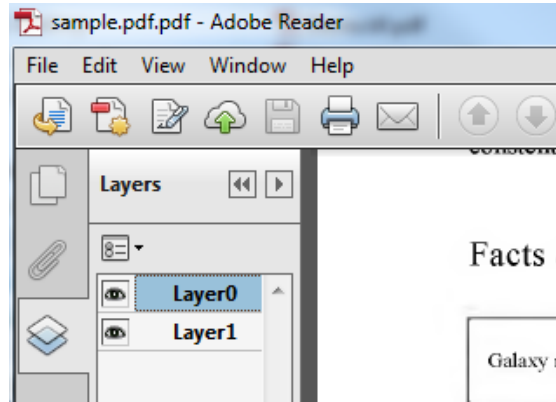
Now it is possible to specify the color used for filling the areas manually with **IPrepareImageMode::BackgroundFillingColor** property:



The default value of this property is -1, which means that the color is determined by ABBYY FineReader Engine automatically.

POSSIBILITY TO SEE PDF LAYERS IN PDF VIEWERS

A new property `IPDFMRCParams::AssignPdfLayersToMrcPlanes` was added in this release. This property manages the availability of PDF layers in the output PDF file. If you set it to `TRUE`, PDF layers are assigned to MRC image planes, so that in some PDF viewers such as Adobe Acrobat you can choose which layers to view:



This property is only taken into account for image-only and image-on-text PDF files.

The default value of this property is `FALSE`.

NEW METHOD `IImageDocument::RemoveColorObjectSEX`

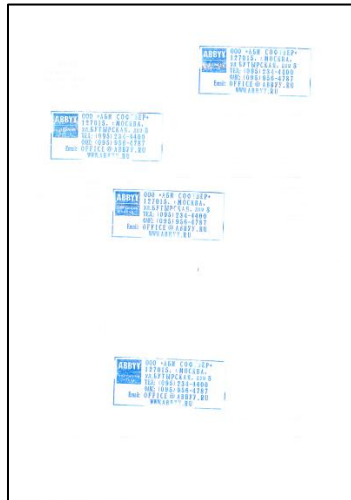
This feature can be useful for documents with black text and white background that contain colored elements that don't need to be recognized. For example, for monochromic documents with colored stamps only black-and-white layer can be recognized and the colored elements can be added to output document.

`IImageDocument::RemoveColorObjectsEx` method allows you to remove color objects of specified hues from original document

P.O. NO	TERMS	REF	SHIP DATE	SHIP VIA	FOB	PROJECT
1505	Net 30	618004	06/04	Carrier		
QTY	ITEM	DESCRIPTION	RATE	AMOUNT		
7		49918 - Gloves or mittens, NCK, in boxes	150.00	1050.00		
42		15583 - Bulk, NCK, inflated	275.00	11550.00		
24		15520 - Athletic or Sporting Goods, NCK	317.00	7608.00		
				TOTAL	20208.00	

P.O. NO	TERMS	REF	SHIP DATE	SHIP VIA	FOB	PROJECT
1505	Net 30	618004	06/04	Carrier		
QTY	ITEM	DESCRIPTION	RATE	AMOUNT		
7		49918 - Gloves or mittens, NCK, in boxes	150.00	1050.00		
42		15583 - Bulk, NCK, inflated	275.00	11550.00		
24		15520 - Athletic or Sporting Goods, NCK	317.00	7608.00		
				TOTAL	20208.00	

The hues of removed objects are set by HSL representation. The method allows replacing removed objects with the specified color. It also allows saving a separate image containing only the extracted objects:



NEW ADVANCED LANGUAGE DETECTION MODE

New property **RecognizerParams::LanguageDetectionMode** was added on this release. This property will be useful for recognition of documents the language of which is not known to you. If language detection mode is on, the recognition languages are selected from the list of languages specified in the **TextLanguage** property. Language detection was significantly improved from the previous release and now it turns on automatically if it is necessary.

This property has three modes: TSPV_Auto, TSPV_No and TSPV_Yes. The default value of this property is TSPV_Auto. In this mode ABBYY FineReader Engine will automatically determine if this processing mode should be used, depending on the situation.

Auto detection will be useful only if TextLanguage set contains a combination of CJK and European languages. New feature will increase the speed and improve the quality of recognition for documents with European languages.

In the scenario when you know that all languages, specified in TextLanguage set, are present in the document, that you process, we recommend setting RecognizerParams::LanguageDetectionMode to TSPV_No.

Old property RecognizerParams::DetectLanguage is marked as deprecated.

NEW PREDEFINED FILTER FNF_PDF IN FONTNAMESFILTERSENUM

FontNamesEnum enumeration constants describe predefined filters of font family names. These filters specify the set of fonts to be used during document synthesis. A new value **FNF_PDF** was added to FontNamesFiltersEnum. If this filter is applied, document synthesis uses font families the names of which are specified in the resources of the input PDF file. However, the fonts themselves are not extracted from the PDF file; they need to be installed on the workstation to be used.

SKEW CORRECTION IN PRPROCESS STAGE

Now it is possible correct skew at preprocessing stage instead of during image opening.

CorrectSkew property was added to PagePreprocessingParams object. The type of skew correction is defined by the **CorrectSkewMode** property.

ADVANCED ISEMPTYEX METHOD FOR CHECKING IF THE PAGE IS EMPTY

IsEmptyEx is a new method which allows you to specify additional parameters during empty page detection.

With **EmptyPageDetectionParams** object you can define the number of letters and text objects that a page can contain and still be considered empty. It is possible to set maximum black percentage and to specify if the page must be searched for barcodes. You can also set the page rectangle, so that any garbage on the margins does not affect the result.

A method **IEngine::CreateEmptyPageDetectionParams** was included in API to create the EmptyPageDetectionParams object.

POSSIBILITY TO EXPORT TIFF FILES WITH ONE STRIP

A strip is a subsection of the image composed of one or more rows. A TIFF image may be composed of one or more strips. Now it is possible to deliberately produce TIFF file composed of one strip by setting **ITiffExtendedParams::WriteSingleStrip** to TRUE. The parameters for TIFF files producing can be used Image::WriteToFile method.

That feature was implemented as some archive writers don't support valid TIFF files with several strips.

NEW PROPERTY RECOGNITION SET

A property **ITextLanguage::RecognitionSet** returns the full letter set used for recognition with this **TextLanguage**, combining all letter sets of its base languages and additional letter sets.

CORRECTION DISPLAY OF PDF FILES WITH PMINGLIU/MINGLIU FONTS IN PDF VIEWERS

Now all exported PDF files with PMingLIU and MingLIU fonts are displayed correctly in all PDF viewers. Previously there were problems with display due to errors of some viewers. As a workaround these fonts are now embedded to the output PDF file.

Please note that if for a Chinese-language document the PMingLiU/MingLiU font is used, it will be embedded into output PDF file regardless of the value of the property IPDFExportFeatures::FontEmbeddingMode.

INFORMATION ABOUT ADDING COMMENTS IN PROFILES IN HELP FILE

The information and an example on how to add comments, were in the section Working with Profiles Comments can be added by starting a line with a semicolon.

POSSIBILITY TO ENABLE AND DISABLE INTERPOLATION IN PDF VIEWERS

Interpolation in PDF viewers can insignificantly affect the visual quality of PDF file. A new property **IPDFPictureCompressionParams::EnableInterpolationMode** allows to disable interpolation in PDF Viewers. The property has three modes: TSPV_Yes, SPV_No and TSPV_Auto. Note that if the IPDFExportParams::PDFAComplianceMode is set to PDF/A, interpolation will be always disabled as it is required by specification of PDF/A format.

The default value of this property is TSPV_Auto, which means that the interpolation will be turned off for PDF/A-compliant formats and on otherwise.

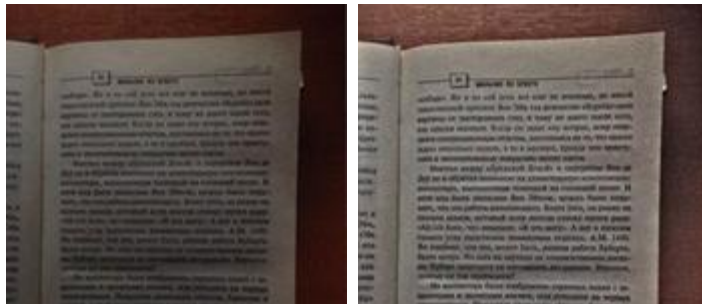
NEW PROPERTY IENGINE::AVAILABLEPREDEFINEDLANGUAGES

New property **IEngine::AvailablePredefinedLanguages** returns the collection of predefined languages that are available under the current license.

NEW PROPERTY FOR SHADOWS AND HIGHLIGHTS CORRECTION IN PHOTOGRAPHS

IPagePreprocessingParams::CorrectShadowsAndHighlights property was added in this release. This property allows correcting excessive shadows and highlighting during image preprocessing. This property is designed for use with photographs only.

Three states are available: TSPV_Yes, TSPV_No and TSPV_Auto. By default the property is set to TSPV_Auto.



Before

After

NEW PROPERTY IHTMLEXPORTPARAMS::USECSS

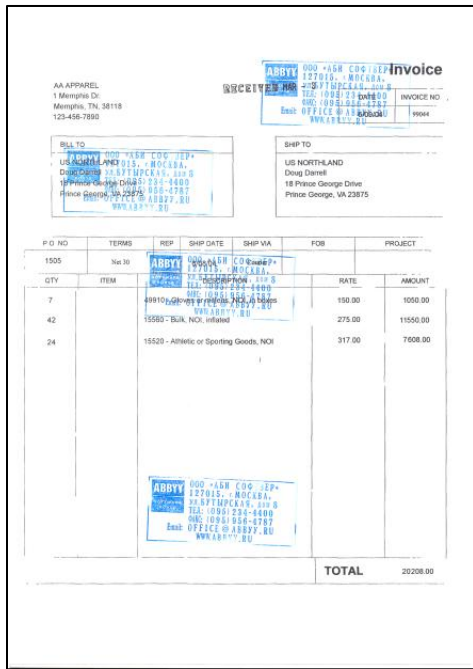
This property determines if a separate style sheet file (.css) is created. In previous releases style sheet file was always created. Now it is possible to use built-in style sheet file by setting IHTMLExportParams::UseCss to FALSE.

The default value of this property is TRUE.

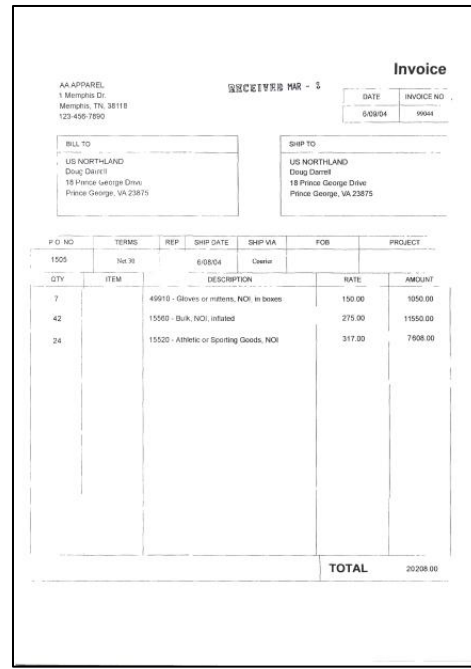
POSSIBILITY TO IMPROVE RECOGNITION QUALITY BY REMOVING COLOR OBJECTS BEFORE RECOGNITION DURING PREPROCESSING STAGE

This feature can be useful for documents with black text and white background that contain colored elements (for example, stamps) that don't need to be recognized.

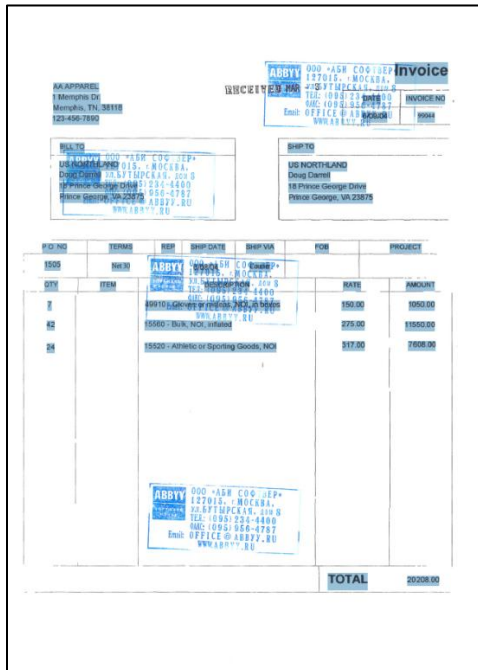
New property **IPageProcessingParams::ProhibitColorObjectsAtProcessing** allows filtering out color objects on the image before layout analysis and recognition. After processing is complete the color objects can be put back on the image again:



1.INPUT DOCUMENT CONTAINS COLOR OBJECTS DURING PROCESSING



2.COLOR OBJECTS ARE IGNORED



3. Color objects are restored in export result

If this property is set to FALSE, the **ColorObjectsProhibitingParams** property is ignored. The default value of this property is FALSE.

A new **ColorObjectsProhibitingParams** object has been added. It is used for tuning parameters of filtering out the color objects on the image before starting processing. This kind of preprocessing can be useful in cases when the document to be recognized has color stamps, signatures etc., which can reduce recognition quality. The parameters are only taken into account if the **ProhibitColorObjectsAtProcessing** property is set to TRUE.

It is possible to specify in properties of this object the following parameters:

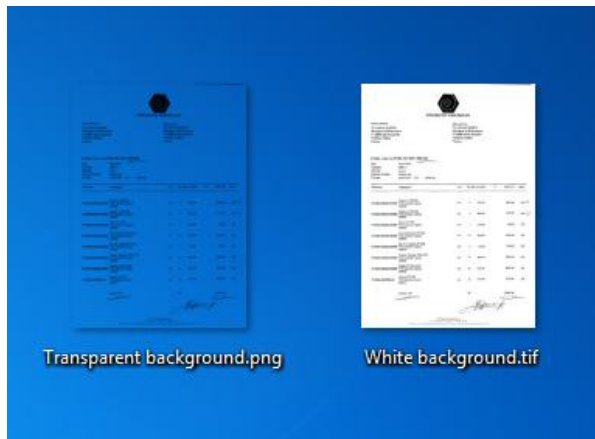
- the color which must replace the removed color objects
- the collection of the hues of the objects which must be filtered, in HSL representation.
- if, after processing is complete, the color objects must be put back on the image again.

A new property **IPageProcessingParams::get_ColorObjectsProhibitingParams** returns a **ColorObjectsProhibitingParams** object.

Another way to improve recognition quality by removing color objects is to use **ImageDocument::RemoveColorObjectsEx** method, added in previous release, but it doesn't allow restoring color objects on the image again before export.

POSSIBILITY TO REPLACE BLACK OR WHITE COLOR OF EXPORTED PNG IMAGES WITH TRANSPARENT

From now on it is possible to replace black or white color of exported PNG images with transparent:



A new object **PngExtendedParams** was added. This object provides functionality for tuning the parameters of saving a black-and-white image to PNG format (IFF_Png format) using the **Image::WriteToFile** method.

To replace the color with transparent it necessary to specify the color in **IPngExtendedParams::TransparentColor** property. Only black and white colors are currently supported.

The default value of this property is -1, which means that no color will be replaced with transparent color.

NEW METHOD IENGINE::CREATEMULTIPAGEIMAGEWRITEREX

The method `IEngine::CreateMultipageImageWriterEx` allows to set extended parameters of image saving (`JpegExtendedParams`, `PngExtendedParams` or `TiffExtendedParams`) when creating the `MultipageImageWriter` object.

NEW ATTRIBUTE 'ROTATION' IN XML EXPORT SCHEME

New attribute 'rotation' in 'page' tag defines the type of rotation applied to original page image before processing. It can have one of the following values: `Normal`, `RotatedClockwise`, `RotatedUpsideDown`, `RotatedCounterclockwise`.

NEW PROPERTY IRTFEXPORTPARAMS::KEEPPAGECREAKS

`IRTFExportParams::KeepPageBreaks` property specifies if the page breaks must be retained in the output RTF document.

NEW PROFILE FOR SPEED UP OF BARCODE RECOGNITION

New profile `BarcodeRecognition_Speed` was added in this release. It enables fast barcodes extraction.

New `BarcodeRecognition_Accuracy` profile is equivalent to the `BarcodeRecognition` profile available in previous releases.

In some cases `BarcodeRecognition_Speed` speeds up the processing almost by 2 times. However, it can decrease the recognition quality. Below you can see the changes in results of processing with the `BarcodeRecognition_Speed` profile compared to the `BarcodeRecognition_Accuracy` profile for some types of barcodes.

Barcode type	Speed up in <code>BarcodeRecognition_Speed</code> profile	Lost in quality in <code>BarcodeRecognition_Speed</code> profile
1D Barcodes, Code 128	14%	10%
1D Barcodes, Code 93	73%	0%
1D Barcodes, EAN 8	29%	12%
1D Barcodes, UPC-E	116%	0%
2D barcodes, Aztec	59%	50%

2D barcodes, MaxiCode	94%	37%
--------------------------	-----	-----

Tests were run on the machine with following configuration: CPU - Intel(R) Core(TM) i5-3450 CPU @ 3.10GHz, Memory - 2.0 Gb.

NEW SECTION IN DEVELOPER'S HELP "PREDEFINED PROFILES SPECIFICATION"

The new section contains a full list of all settings used by ABBYY FineReader Engine predefined profiles.

SUPPORT OF CORRUPTED TIFF FILES OPENING

Starting with this release, FineReader Engine can open corrupted TIFF files. It can be done only if one last line is corrupted, otherwise an error occurs and the corrupted file is not opened.

If the TIFF image which has been opened is corrupted, a warning message containing the number of the damaged page in the image file will be shown.

EXPORTING ALTO UP TO VERSION 3.0

It is possible to export OCR data into ALTO format according to the following ALTO standard versions:

- 2.0
- 2.1
- 3.0

The default version is 2.0, same as it was in the previous release.

MEMORY SAVING IN CASE OF NO NEED IN COORDINATES ON ORIGINAL IMAGE

Instruction `IPrepareImageMode::KeepOriginalCoordinatesInfo = false` tells the Engine not keeping image transformation matrix required for backward object coordinates calculation used for getting coordinates on original (source) image.

If there is no need in keeping the transformation data, it helps to preserve storage and memory space. For example, during processing of b/w images (~100Kb) the transformation data could be up to 1Mb per image.

CROP FUNCTION SUPPORTS GREYSCALE AND B/W IMAGES

Starting from this release it is possible to run `ImageDocument::CropImage()` function on greyscale and black and white images. Previously that was possible only for color images.

FINE TUNING OF PAPER SIZE DETECTION

The Engine is able to detect paper boundaries on a scan automatically (`PageAnalysisParams::NoShadowsMode = false`, default value).

This helps to rid of garbage detection as text near borders of a scan. In some cases, this detector can fail. For example, if a scanned document contains a big dark picture it might be taken as a scanning background and removed from a document area as a scanning shadow.

To prevent such mistakes a host application may advise the Engine to limit the scanning shadow detector hypotheses by providing information on what part a source document occupies on a scan.

API provides PageAnalysisParams::PaperSizeDetectionMode for this purpose.

IMPROVED MRC QUALITY

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

NEW PROPERTY IENGINE::VERSION

We improved all predefined MRC modes

The main target was to improve visual quality because predefined MRC modes available in the previous releases provide too low quality and our customers sent us many complaints regarding that.

MINSIZE PREDEFINED MODE.

In exchange of 32-38% increase of a file size, you get 6-15% faster speed and significantly better document image quality:

Old settings	New settings
	
	



等技術の防毒軟件，就可通過大部分的考驗。而繼紅色代碼之後，Nimda 利用同樣方式導致網絡通信嚴重異常。更甚的是，2003

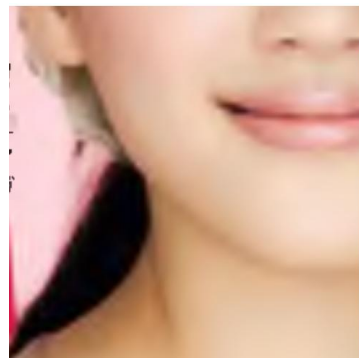
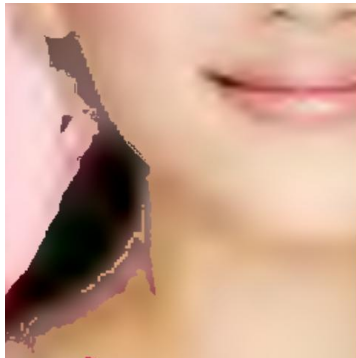
等技術的防毒軟件，就可通過大部分的考驗。而繼紅色代碼之後，Nimda 利用同樣方式導致網絡通信嚴重異常。更甚的是，2003

モジュラーベルト&ハーネス	35-36
ノートパソコンとカメラのコンビ	37-40
ショルダーバッグ	41-45
スリングバッグ	46

モジュラーベルト&ハーネス	35-36
ノートパソコンとカメラのコンビ	37-40
ショルダーバッグ	41-45
スリングバッグ	46

심리 다룬

심리를 다룬



IPCC第4次報告書より
21世紀末における世界

IPCC第4次報告書より
21世紀末における世界



MAXQUALITY PREDEFINED MODE

In exchange of negligible image size increase (~0%), new settings provide 2-11% faster speed and better visual quality for a document image:

Old settings	New settings
 <p>手工漢堡包 556 士香濃加上手工嚴擇包的配 選製是一顆美味的享受。</p>	 <p>手工漢堡包 556 士香濃加上手工嚴擇包的配 選製是一顆美味的享受。</p>
<p>验。而继红色代码之后，Nimda 利用同样方式导致网络通信严重异常。更甚的是，2003</p>	<p>验。而继红色代码之后，Nimda 利用同样方式导致网络通信严重异常。更甚的是，2003</p>
 <p>You may argue that the Dell a operators were enhanced Load bearing ved's, which they cut away the ched ammo pouches</p>	 <p>You may argue that the Dell a operators were enhanced Load bearing ved's, which they cut away the ched ammo pouches</p>
<p>2 With kitchen shears, cut off leaf tips.</p> <p>3 Using a melon baller,</p>	<p>2 With kitchen shears, cut off leaf tips.</p> <p>3 Using a melon baller,</p>
<p>Adobe</p>	<p>Adobe</p>

<h1>الجيدة،</h1>	<h1>الجيدة</h1>
<h2>تحسُن وسائل</h2>	<h2>تحسُن وسائل</h2>
<p>Panopolis及Millie's Cookies 原價每盒HK\$100，滿額可作HK\$20現金使用。 *請於滿額時注意此券，每次只可善用一張 *此券只適用於指定連鎖咖啡廳分店 *此券不能兌換現金或作其他用途及 不可與其他優惠共同使用 *Sales Service Partner Hong Kong Ltd. 保留更改及使用 此優惠之最終權 *優惠期至 2010年8月30日 查詢電話：2520 4141</p>	<p>Panopolis及Millie's Cookies 原價每盒HK\$100，滿額可作HK\$20現金使用。 *請於滿額時注意此券，每次只可善用一張 *此券只適用於指定連鎖咖啡廳分店 *此券不能兌換現金或作其他用途及 不可與其他優惠共同使用 *Sales Service Partner Hong Kong Ltd. 保留更改及使用 此優惠之最終權 *優惠期至 2010年8月30日 查詢電話：2520 4141</p>
<p>'파리의 여자들에게선 뒤를 돌아보게 같지도 않고 진한 향수 냄새를 풍기는 투, 표정, 체형, 사고방식 등이 패션 이</p>	<p>'파리의 여자들에게선 뒤를 돌아보게 같지도 않고 진한 향수 냄새를 풍기는 투, 표정, 체형, 사고방식 등이 패션 이</p>

BALANCED PREDEFINED MODE

In exchange of 14-15% increase of a file size, you get 6-15% speed increase and significantly better document image quality:

Old settings	New settings
	
<p>are prettly damn hard impo... ble in the UK, in the states they're g</p>	<p>are prettly damn hard impo... ble in the UK, in the states they're g</p>

<p>株式会社 商品市況研究所 http://www.cmri.jp/</p>	<p>株式会社 商品市況研究所 http://www.cmri.jp/</p>
<p>モジュラーベルト&ハーネス 35-36 ノートパソコンとカメラのコンビ 37-40 ショルダーバッグ 41-45 スリングバッグ 46</p>	<p>モジュラーベルト&ハーネス 35-36 ノートパソコンとカメラのコンビ 37-40 ショルダーバッグ 41-45 スリングバッグ 46</p>
 <p>手工肉排包 556 芝士香濃加上手工肉排包的陪 選製區一體美味的享受。</p>	 <p>手工肉排包 556 芝士香濃加上手工肉排包的陪 選製區一體美味的享受。</p>
<p>اللجوء ريثما</p>	<p>اللجوء ريثما</p>
<p>大力水手飾圖案標 564 三件炸雞套餐，以鬆製 過，雞肉鬆鬆滑滑美味； 薯片由伊威爾合法會製 精心選擇，保證可美可口</p>	<p>大力水手飾圖案標 564 三件炸雞套餐，以鬆製 過，雞肉鬆鬆滑滑美味； 薯片由伊威爾合法會製 精心選擇，保證可美可口</p>
<p>Using a serrated knife, cut off the stem, and trim two inches off the top.</p>	<p>Using a serrated knife, cut off the stem, and trim two inches off the top.</p>
<p>Nowhere's where things start to get a bit more expensive in the real World. Your</p>	<p>Nowhere's where things start to get a bit more expensive in the real World. Your</p>

MAXSPEED PREDEFINED MODE

In exchange of almost no file size increase (2-8%) new settings provide better image quality (see modes above) and noticeably faster processing speed (30-45%).

CONVENIENT METHOD FOR GETTING RECOGNIZED WORD REGION

Many customers work with recognized words instead of recognized characters. In most of the cases, such customers need to highlight a word of matter in graphic user interface. The highlighting requires a word-bounding rectangle (region).

Since words may include characters of different height and could span several lines the task becomes non-trivial. To make it easy and facilitate our customers we added new property to IWord interface that returns a word-bounding region – IWord::Region.

LARGE DOCUMENT CONVERSION TO SEARCHABLE PDFS AT FULL THROTTLE

Super multi-page document conversion to searchable PDF might be a challenging task if your PC/Server has limited resources. This is a resource-intensive task and at certain points becomes a very long time-consuming one.

Such tasks and issues are common for mid and big MFPs where one can put a big pile of papers and order a searchable PDF output. A similar scenario you can see in the case of an archive conversion while the archive stores super-multi-page 'files'.

To provide a proper tool we added a new API, and found the best usage scenario for FRE 11 and are introducing this in the R6 release product documentation: look for the article 'Exporting Large Documents'.

The main advantages of the new usage scenario

- dramatic speed increase when exporting large documents
- less RAM resources required
- convenient error handling without losing all the exported data. In case of simultaneous export a single error may cause failure of the whole export stage. If you are processing a large number of pages, the restarting of export can take plenty of time. Now using the new functionality you do not have to stop processing. The errors can be caught and handled inside small portions. So export will be finished faster even if some errors occur during the processing.

Recommendations

- Using the new export mode is reasonable for documents which contain 50 pages or more. For 500 pages and more using ExportFileWriter is highly recommended.
- Export a fixed number of pages at a time. You will need to do some experimenting to pick the best number of pages for your documents. During internal ABBYY tests the portion size of 30 pages was found best for generic documents.

JAPANESE OCR IMPROVEMENTS

Last year R&D and ABBYY Japan spent much time jointly in the project for improving Japanese OCR. The joint team identified and solved many issues. The project is not finished, but FRE 11 R6 contains intermediate results including the following improvements:

- Old Kanji characters issue. ABBYY Japanese OCR is capable of recognizing Kanji characters not of frequent use nowadays. The solution: JapaneseModern is a new OCR language available in the product. It does not include outdated Kanji. Our customers should use it to recognize modern documents with the best accuracy.
- Recognizing Katakana as Kanji (Hiragana) Character Issue (タ ->夕, オ->才, エ->工, etc.). ABBYY OCR mixes up similar characters where a human can easily identify correct ones. The solution: application of
 - Context-based heuristics. It is something like 'recognize " 1 -" instead of " 卜" because a string looks like an address'.
 - Japanese orthography rules. It is something like 'do not put a Hiragana character in between two Katakana characters'.

- Missing characters issue. The solution: few frequent and missing modern Japanese characters were trained and added into the recognition set.
- Lack of Japanese text models issue. The solution: new models (helps with similar characters and mixed language recognition) are added:
 - Date. Such as 2014 年 10 月 19 日 (日) .
 - Address. Such as 東京都港区赤坂 2 丁目 1 7 - 2 2 .
- Dictionary issue. Existing dictionary contained 'garbage' and misses some initial forms. The solution: incorrect entries were removed, new initial forms are added (borrowed from UniDic dictionary popular in Japan).

FRE 11 R6 leverages differently on different test-beds: sometimes the improvements give nothing, sometimes the effect is very small (1% less errors), sometimes it is worth to mention (24.6% less errors).

ARABIC FATHATAN (َ) DIACRITICS SUPPORT

Arabic OCR got support for the Arabic Fathatan diacritics recognition. The initial request came from a ABBYY FlexiCapture 11 user. Since this improvement increases overall Arabic OCR accuracy, we decided to add this feature to FRE 11 R6.

Please notice that the improvement is valid for the last symbol in a word:



THAI OCR ACCURACY IMPROVEMENT

In a scope of entering into VIP account, we faced with low Thai OCR accuracy on several font types comparing to competitors.

The font types are following:

- Tahoma
- TH SarabunPSK
- TH Niramit AS
- TH Kodchasal
- TH Baijam
- TH Fah Kwang

Average accuracy was like this according to the customer's report:

Engine	Competitor A	Competitor B	ABBYY SDK
Average accuracy	78.5%	93.1%	80.7%

Dedicated support of colleagues from ABBYY Japan and R&D professionals' efforts make it possible to overcome competition with the following result:

Engine	Competitor A	Competitor B	OCRT 14
Average accuracy	78.5%	93.1%	91.3%

The improvements in OCRT 14 were backported to OCRT 13 used in FRE 11 R6 release.

Internal test for FRE 11 R6 shows that it has 11%-13% better accuracy in Thai OCR on the test-bed containing 35 images of different document types.

EXTENDED INFORMATION ON CHARACTER POSITION IN A WORD

Previously the `IsWordStart` property tells if a character starts a word. For right-to-left writing, this parameter was ambiguous: logically a word starts from the right end whereas geometrically it starts from the left. To solve the ambiguity API and XML export got two new properties:

- `IsWordLeftmost`. Specifies whether the character is the leftmost character in a word.
- `IsWordFirst`. Specifies whether the character is the first character in a word.

The old property `IsWordStart` is declared deprecated and will be removed from API and XSD in FRE 12.

SMALLER PDF WITH PAGES OF MIXED COLORITY

Not all pages in a document are similar: one may contain color graphics; another may contain b/w text only. Here one can make PDF file smaller: compress color images with higher resolution and less compression, compress b/w image with lower resolution and higher compression. Frankly speaking you may notice that most of PDF creation tools allow you to set up individual resolution and compression for each type of pages: color, grayscale, and black and white.

In one of the previous FRE 11 releases, we added PDF exporting API for setting up individual compression for each page type described above.

Now we present additional API for setting up individual resolution:

- `BwPictureResolutionParams`
- `ColorPictureResolutionParams`
- `GrayPictureFormats`

In combination with existing API for compression selection PDF exporting API lets our customers to save PDF file size same way as other PDF creation tools do.

Please notice that new API does not work in MRC mode. From certain point of view, MRC and individual resolution/compression setting are two different approaches to the same goal: good look with low size. You cannot and should not mix them up; you have to choose one most appropriate for you.

FARSI OCR

From this release Farsi language is supported. It is added to the list of Predefined Languages. Win32 standard language identifier is equal to 1065.

The language is available, if Arabic module is included in license. Please note that as of now, dictionary support for this language is not available.

Internal accuracy measurement and competitive comparison look as follows:

Test-bed	OCR Engine	Error level	
		Characters	Words
Book	FRE 11 R6	9,9%	12,1%
	Verus 4.0	15,8%	11,1%
	ReadIRIS 14	12,7%	16,4%
Doc	FRE 11 R6	8,2%	8,5%
	Verus 4.0	18,5%	11,9%
	ReadIRIS 14	14,2%	13,6%
Wiki 1	FRE 11 R6	11,8%	10,9%
	Verus 4.0	22,6%	12,4%
	ReadIRIS 14	49,9%	16,3%
Wiki 2	FRE 11 R6	21,9%	14,3%
	Verus 4.0	31,2%	17,1%
	ReadIRIS 14	67,8%	47,6%
Wiki 3	FRE 11 R6	13,3%	10,6%
	Verus 4.0	31,9%	21,9%
	ReadIRIS 14	100,0%	100,0%

'BARCODERECOGNITION_ACCURACY' BECAME MORE ACCURATE

The next technology cycle (14/15 version) includes a task for barcode finding improvements following a number of reclamations got from customers using ABBYY products based on 13th technology cycle and earlier. Recently R&D finished the task and back-ported results. That helped to close several high-priority reclamations and improve accuracy results on SDK test-beds like this:

- QRCode: 5.6% less errors
- EAN 13 with supplemental: 20.5% less errors
- EAN 13: 15.38% less errors
- UPC-E with supplemental: 10.5% less errors
- EAN 8 with supplemental: 5.9% less errors

That is in addition to fixed reclamations.

The accuracy improvement is included into 'accuracy' profile only because it brings also a slowdown. The slowdown varies from 5% to 70% depending on image complexity, detection type, and test-bed content (length and specifics).

WARNINGS ARE CONNECTED TO PARTICULAR PAGES

When processing a multi-page document one might get recognition warnings but was not able to identify a document page a warning belonged to. In case of homogenous page images along a document, this does not matter and makes not much sense. 'Homogenous' pages are common for single document scanned. As soon as somebody processes a file (set of different documents) or photos, the situation when a warning raises in the middle page becomes common and a page index starts to be very important.

To address the situation FRE 11 R6 presents updated IFRDocument::IFRDocumentEventsEx interface with extended OnWarning() callback returning PageIndex property.

FRE HELPS IN ZUGFERD SCENARIO

The electronic invoices standard ZUGFeRD was accepted in Germany in 2014. This standard usage is meant to simplify the electronic invoices processing and to increase the efficiency of the work with e-document management.

A ZUGFeRD-compliant e-invoice must combine the human-readable representation and the structured data in XML format that can be recognized by computer.

The file format that satisfies these requirements is the PDF/A-3 format. It is a container for both the visual representation and the structured data in XML format.

ABBYY FineReader 11 Release 3 and later provides the special features for the files in ZUGFeRD standard processing.

To let our customers know FRE 11 is helpful in ZUGFeRD scenario we added the 'ZUGFeRD-compliant electronic invoices' article to the product documentation.

HTTP AND SOCKS 4/5 PROXY SUPPORT FOR A LICENSE ACTIVATION

It is common to use a proxy in a corporate network environment to protect it and manage an employee access to the Internet.

FRE needs to activate a license on a PC to work in case no dongle-based and no pre-activated license is used. If a PC is behind a proxy then FRE protection mechanism needs to know a login and a password for the proxy.

Starting from this release it is possible to specify a proxy type, a login, and a password for it in the 'LicenseSettings.xml' file.

Example:

```
<?xml version="1.0" encoding="utf-8"?>
<LicensingSettings xmlns="http://www.abbyy.com/Protection/LicensingSettings">
  <ProxyServer Type="SOCKS" Host="192.168.56.1" Port="9090" Login="user"
  Password="password" />
  <LocalLicenseServer>
    <ConnectionProtocol ProtocolType="LocalInterprocessCommunication" />
  </LocalLicenseServer>
</LicensingSettings>
```

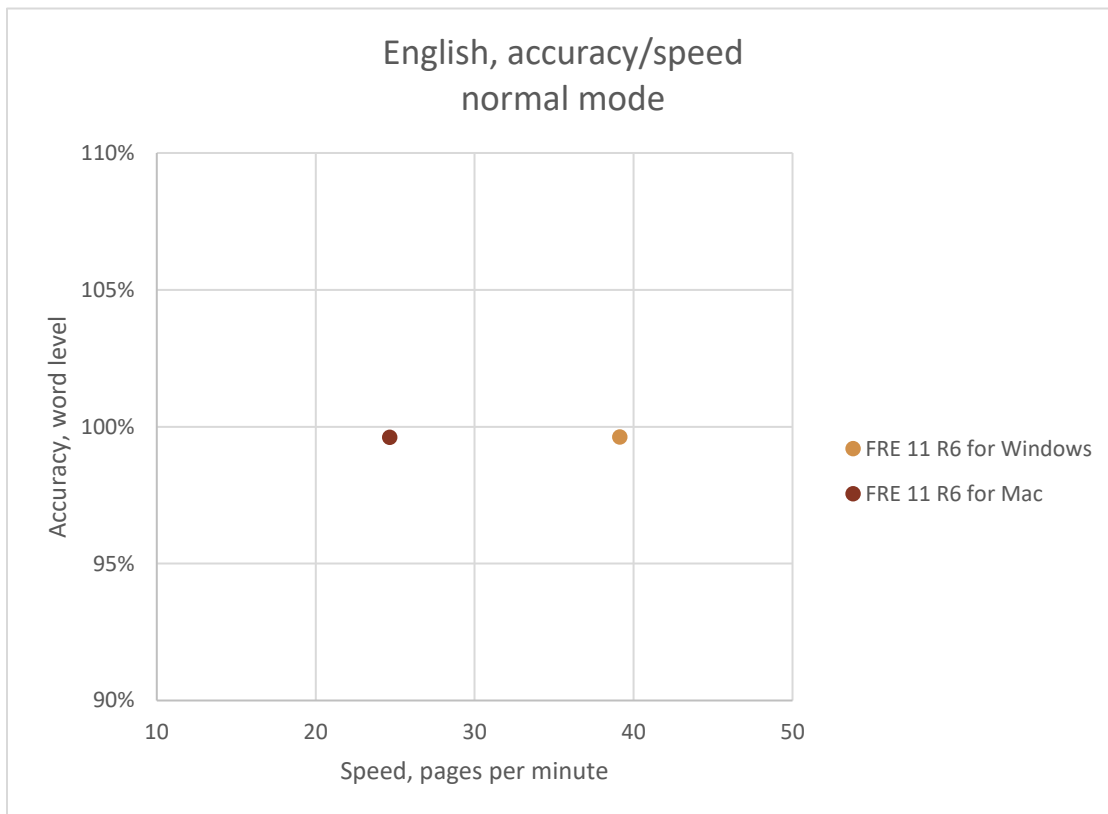
```
<EnableIKeyLicenses Enable="no" />
</LocalLicenseServer>
</LicensingSettings>
```

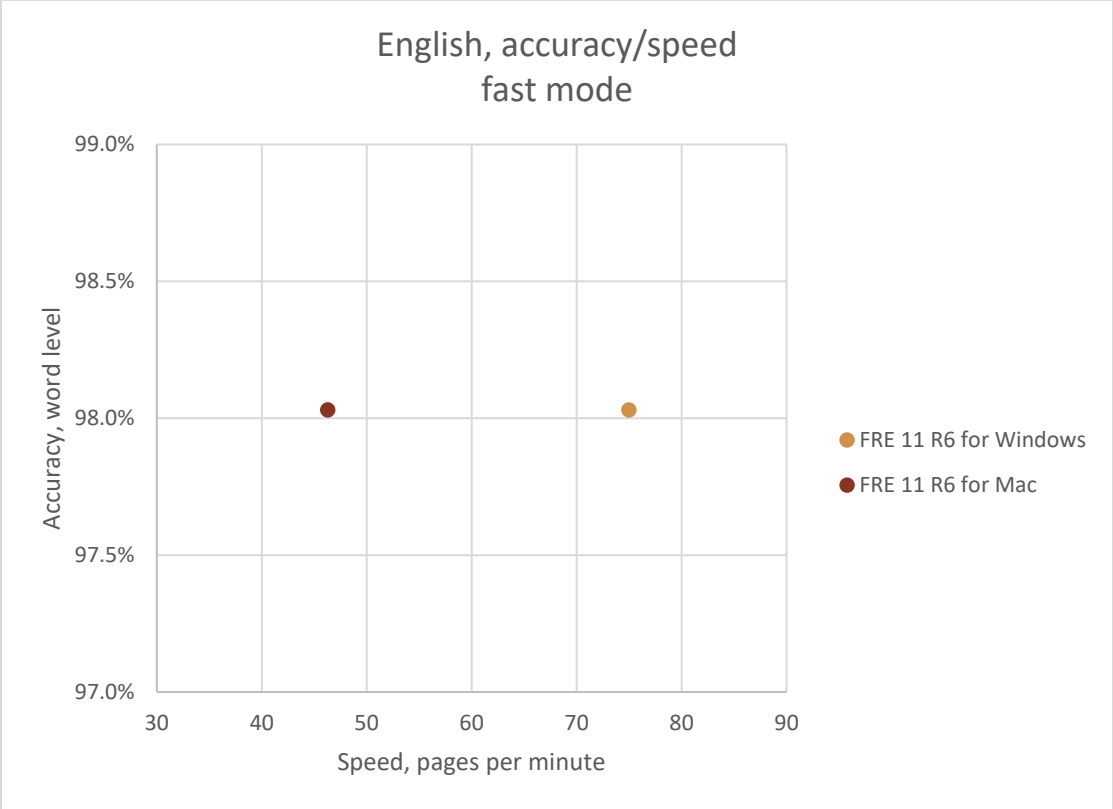
Notice: the product documentation is missing the feature description. The fix is coming in R7.

PERFORMANCE RESULTS

This section contains performance results of the release comparing to the previous releases. The processor of the testing machine is Intel® Core™ i5-3450 CPU (3.10GHz, 4 physical cores) with 8GB of RAM.

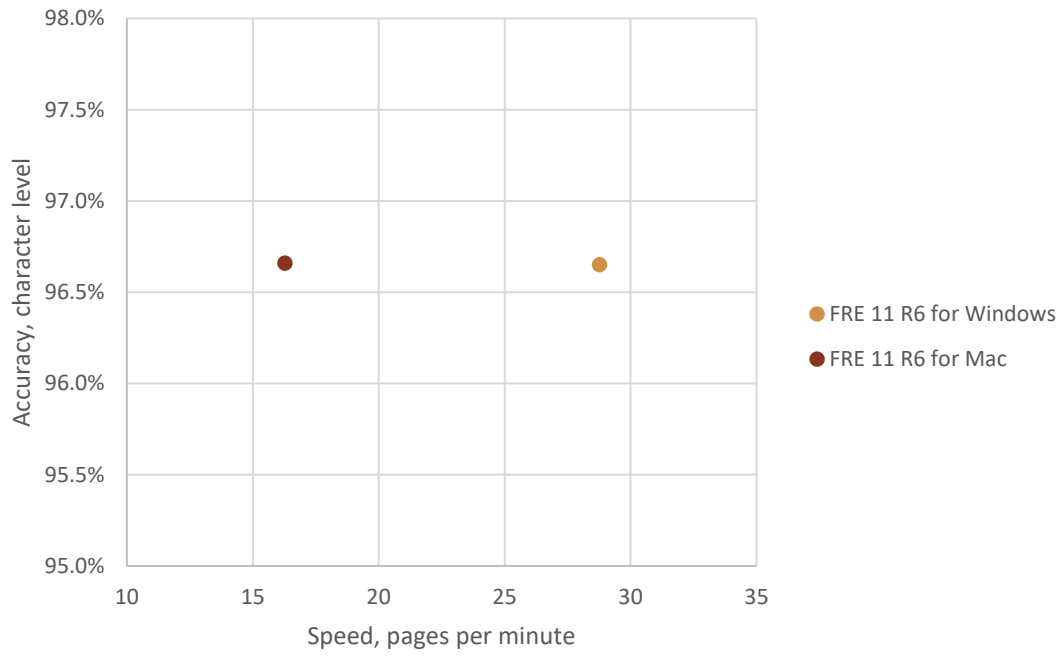
ENGLISH



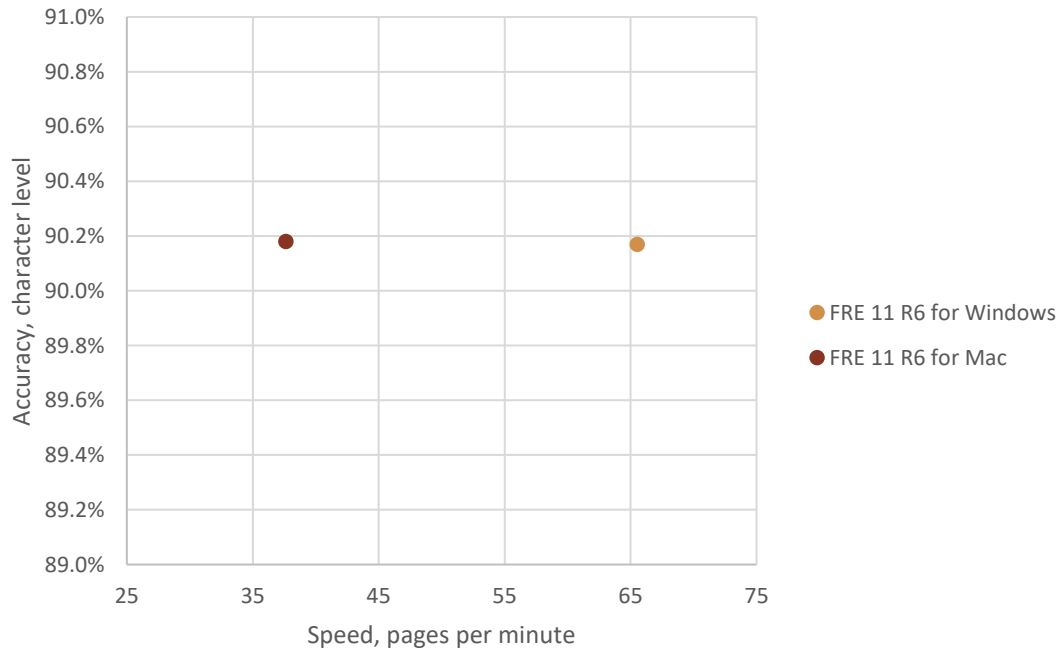


JAPANESE

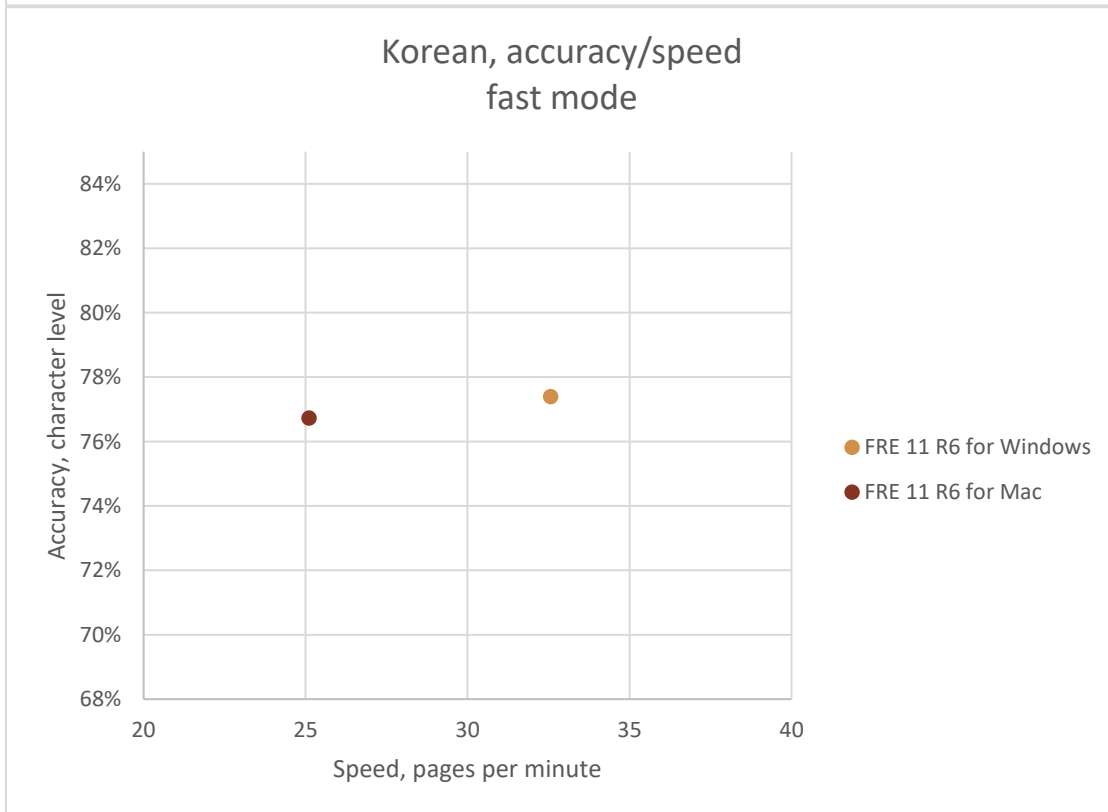
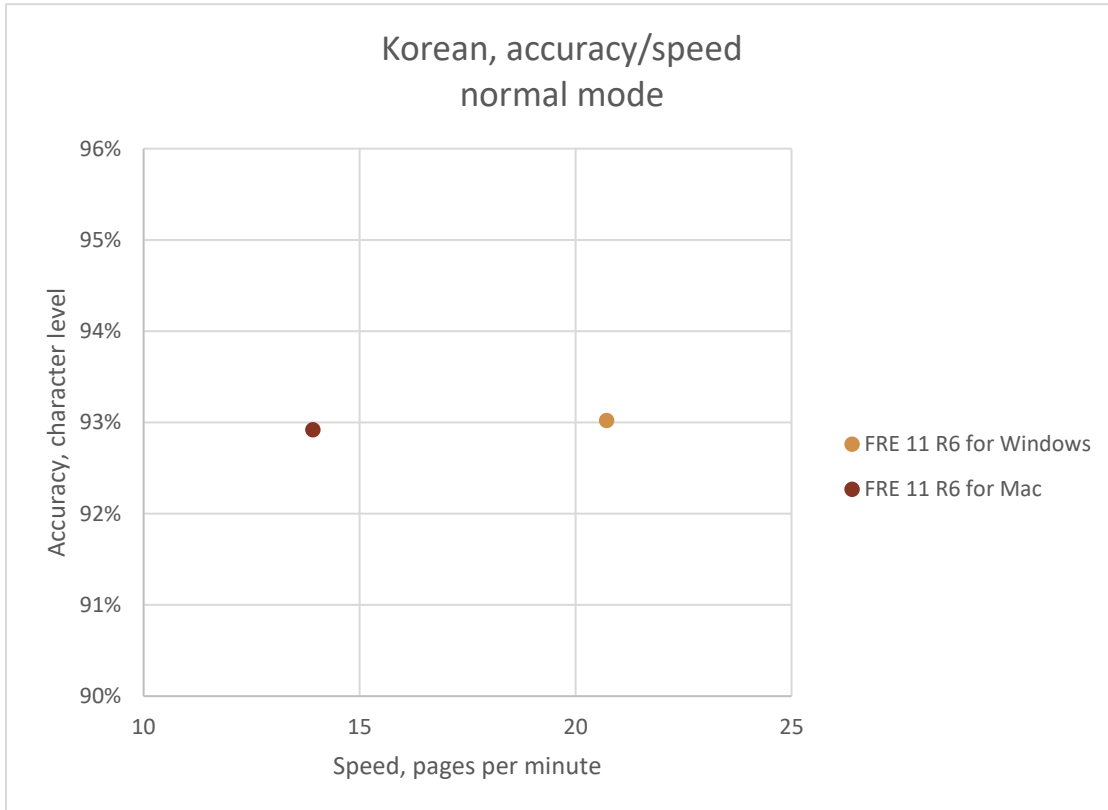
Japanese, accuracy/speed
normal mode



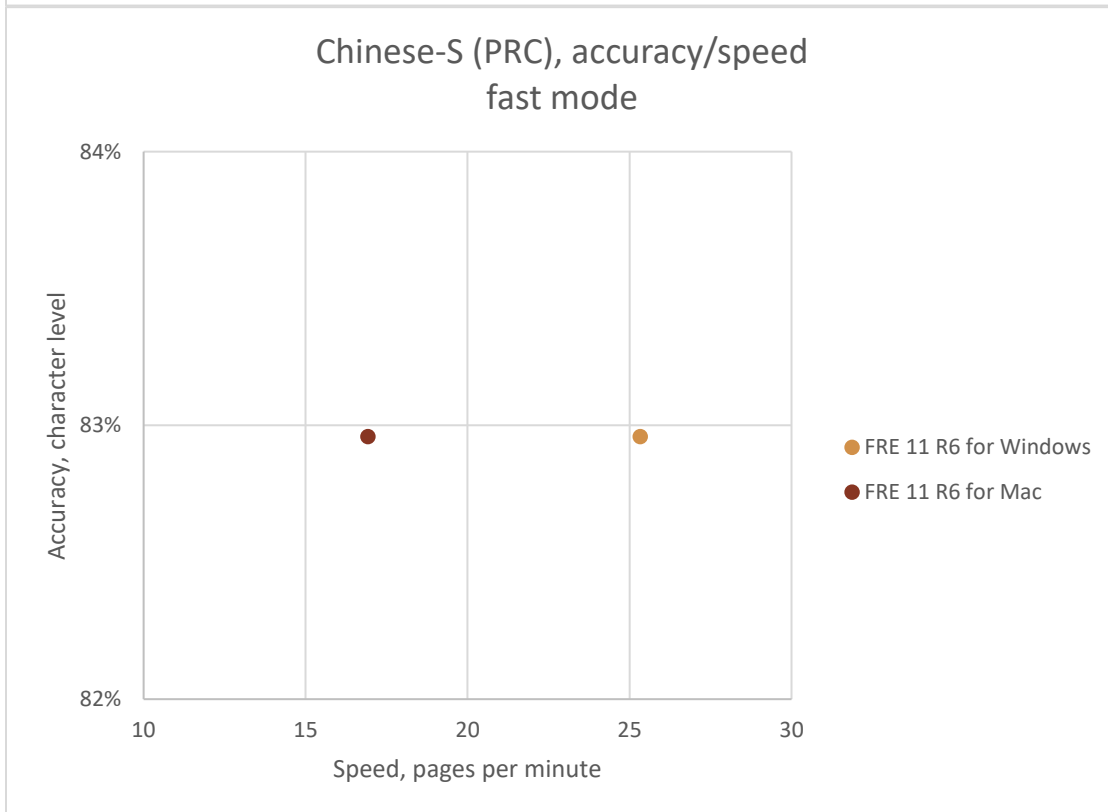
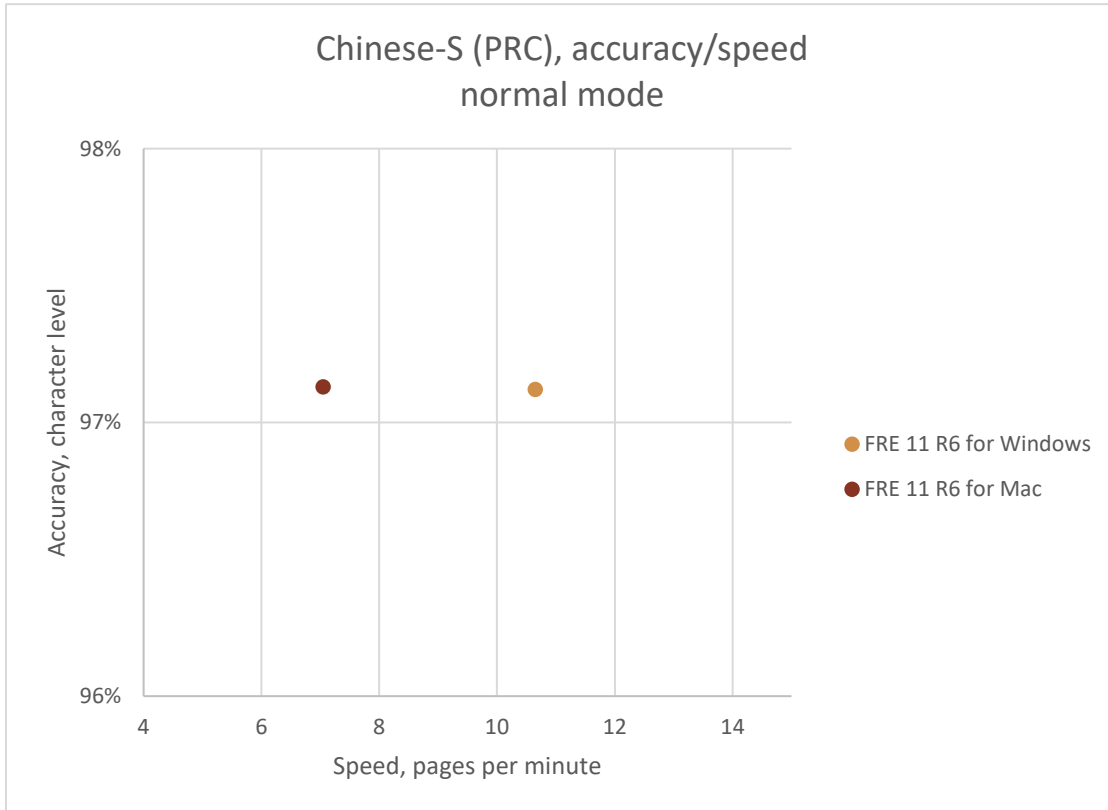
Japanese, accuracy/speed
fast mode

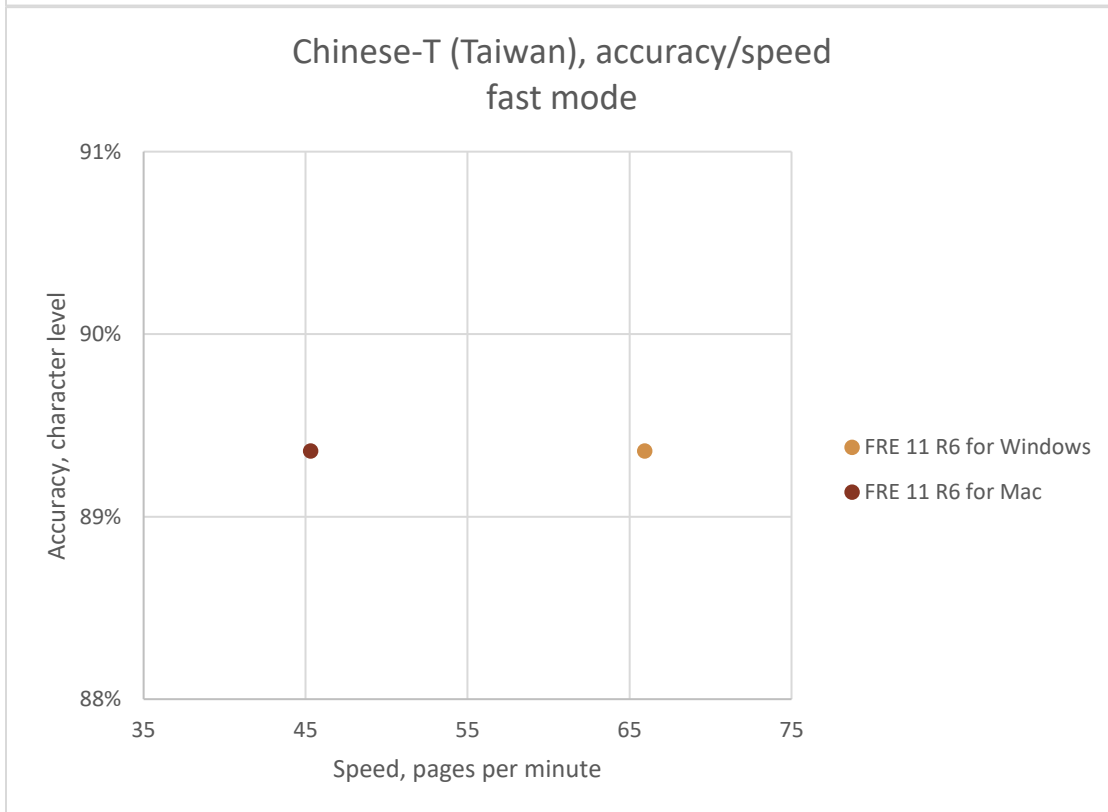
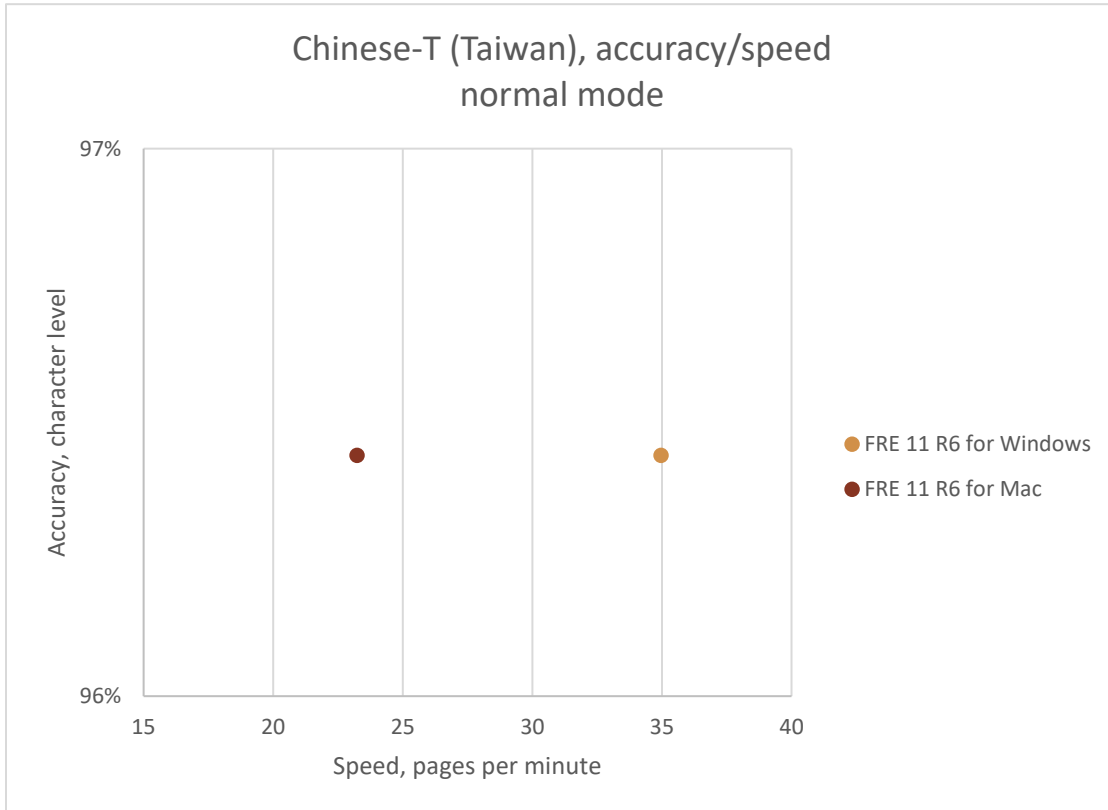


KOREAN

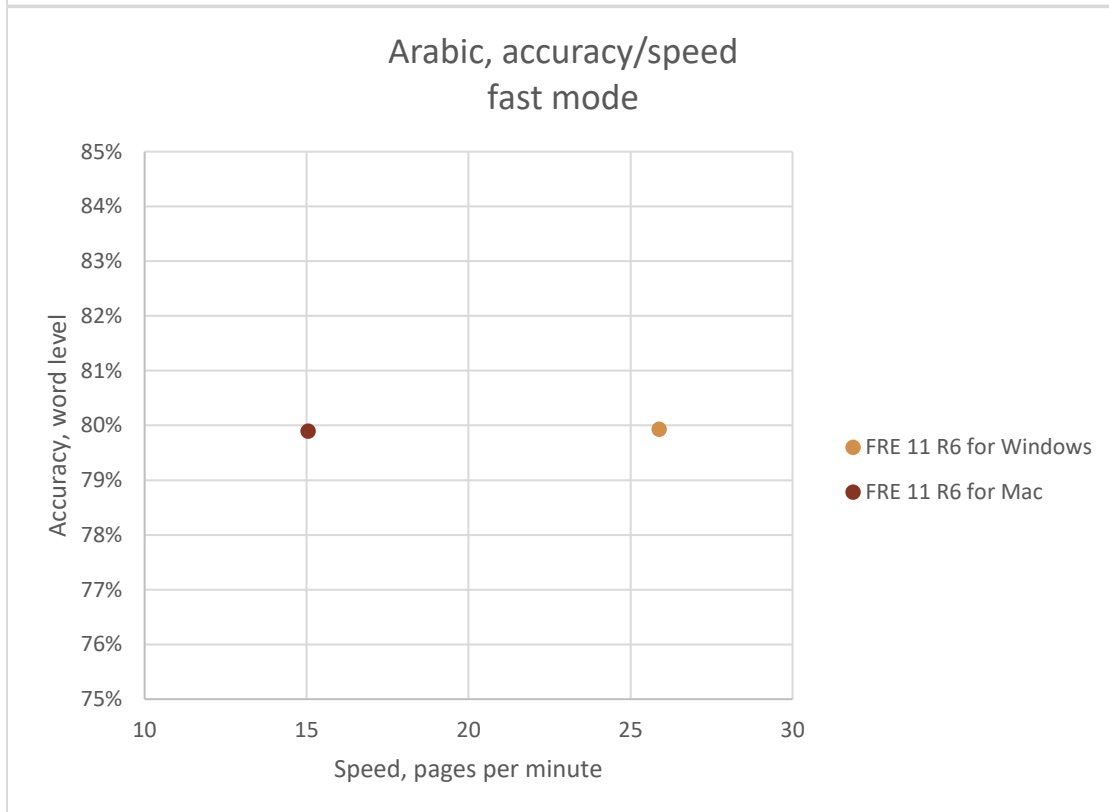
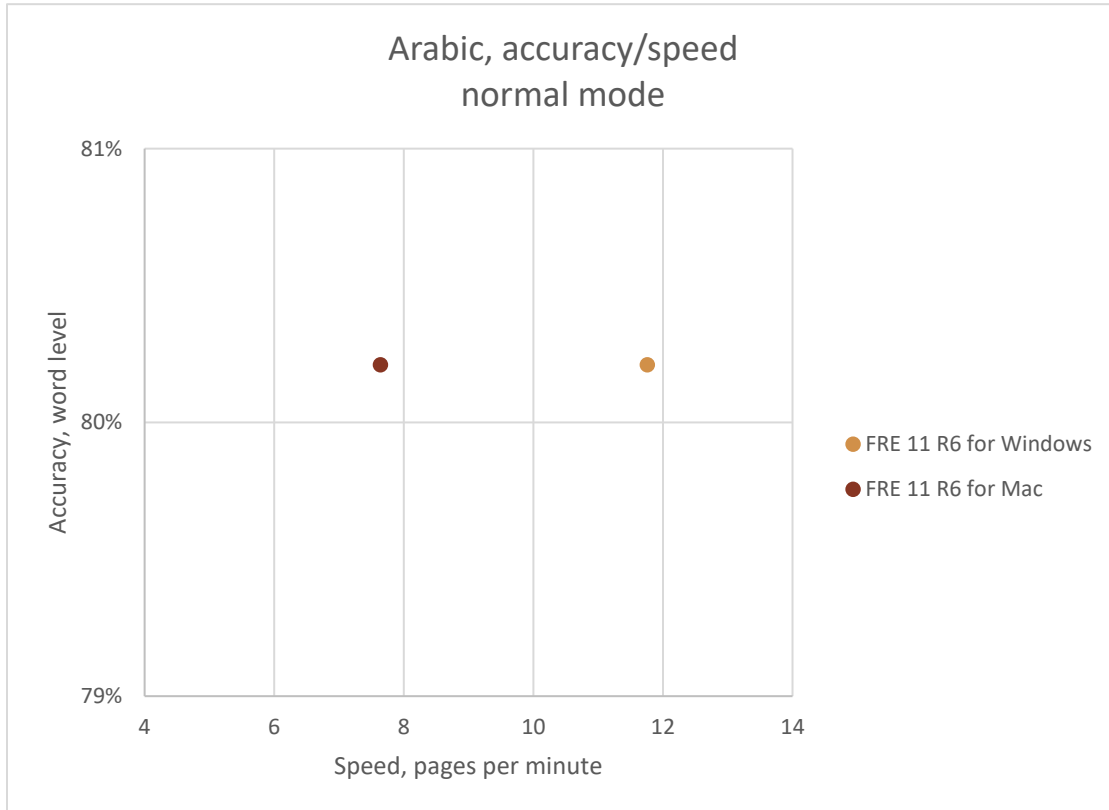


CHINESE PRC





ARABIC



FIXED BUGS

No specific for Mac version bugs were registered and fixed.

KNOWN ISSUES AND WORKAROUNDS

MAC OS VERSION LIMITATIONS

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

- DjVu opening
- Scanning
- ICR/OMR
- Visual Components and other GUI elements
- WDP/WIC/BITMAP input formats and other Windows-specific functionality
- PDF text layer reusing
- Attachments extraction from PDF
- Bookmarks extraction from PDF
- Metadata extraction from PDF (the information about author, keywords, subject and title)

SOME API IS NOT IMPLEMENTED

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

- 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.

FARSI LANGUAGE IS BASED ON ARABIC LANGUAGE

Farsi OCR output indicates Language ID as Arabic (CharParams::LanguageId = LI_ArabicSaudiArabia). This is because Farsi technology preview is based on Arabic OCR language.

The fix will be available in FRE 12.

IPDFMRCPARAMS::MONOCHROMETEXT DOESN'T WORK CORRECTLY

Different algorithms of compressions during export to PDF are used with IPDFMRCPParams::MonochromeText set to FALSE or DEFAULT whereas default value is FALSE.

LANGUAGE AUTO DETECTION USES CJK RESOURCES THOUGH NONE IS SELECTED FOR OCR

'Cjk.*' resource files are used while RecognizerParams::LanguageDetectionMode = TSPV_Yes even though recognition languages set does not include any of CJK languages.

THE PROXY FEATURE DESCRIPTION IS MISSING IN THE DOCUMENTATION

The product documentation is missing a description for the feature 'HTTP and SOCKS 4/5 proxy support for a license activation'.

The will be available in the FRE 11 R7.

30% SLOWDOWN IN ALL SCENARIOS

Internal tests shows that this release got 30% slowdown comparing to R2.

We are investigating the reason and working on a patch.

ADDIMAGEFILEFROMMEMORY DOES NOT OPEN PDF FILES

An attempt to open a PDF file from a memory ends up with the error 'The image file you specify is empty.'

R7 will include a fix.

PDF/A VALIDATION REPORT

The following issues are detected by Adobe Acrobat 11.0.9 (Preflight 11.0.9) for PDF/A files produced by this release:

1. In SourceContentReuseMode = CRM_ContentOnly the Engine sometimes (5% of the test base in Korean language) generates PDF/A-2a and PDF/A-3a files which Adobe Acrobat Preflight validates with the following message "Text is mapped to Unicode Private Use Area but no ActualText entry is present".
2. Sometimes PDF/A-1a/-2a/-3a files are not valid with the following diagnoses:
 - 2.1. "Character references .notdef glyph"
 - 2.2. "Glyphs missing in embedded font".

R6 GM PATCH –

PART NUMBER, BUILD NUMBER

Part#	1161/24
Build#	11.1.14.705064

UPGRADE FROM THE PREVIOUS VERSIONS AND RELEASES

BINARY INCOMPATIBILITY

It is necessary to recompile host application regardless a version of Engine previously used.

API CHANGES

There are no API changes.

CHANGES IN RELEASE 6 UPDATE

ICHARPARAMS::ISWORDSTART PROPERTY IS DEPRECATED

ICharParams::IsWordStart property is deprecated and will be removed in the next major version. *IsWordLeftmost* property must be used instead.

SOME PROPERTIES RELATED TO DOCUMENT STRUCTURE ARE MARKED AS DEPRECATED

The following properties are marked as deprecated:

- *IPageStructure::IsPageOdd*
- *IFootnoteSeries::HasSeparator*
- *IFootnoteSeries::PositionInDocument*
- *IFootnoteSeries::PositionOnPage*
- *IFootnoteSeries::IsNumberingWithSuperscript*
- *ISynthesisParamsForPage::InsertEmptyParagraphsForBigInterlines*

These methods are not implemented and the implementation is not planned.

HFM_FORMAT32 FORMAT IS MARKED AS OBSOLETE

From this release this format is obsolete. Using this constant is equivalent to *HFM_Format40* with the *IHTMLExportParams::UseCss* property set to *FALSE*.

IRECOGNIZERPARAMS.DETECTLANGUAGE PROPERTY IS MARKED AS DEPRECATED

New property *IRecognizerParams::LanguageDetectionMode* is now used instead of *IRecognizerParams::DetectLanguage*. *IRecognizerParams::DetectLanguage* is marked as deprecated.

For purposes of compatibility, new property value is changed whenever you change the *DetectLanguage* property: if you change it to TRUE, the property is set to TSPV_Yes, if you change it to FALSE, this property is changed to TSPV_No.

CREATEIMAGEDOCUMENTSINMEMORY PROPERTY IS MARKED AS DEPRECATED

CreateImageDocumentsInMemory property of Engine Object is marked as deprecated and will be removed in future versions.

FRPAGE::CLASSIFY METHOD INCREASES THE LICENSE COUNTER

In previous release the pages were not counted correctly. Now the license counter is increased after classification. Layouts of classified pages are cleaned if they were not recognized earlier.

CAMERA OCR AND COLOR FILTERING MODULES BECAME PREDEFINED.

New property *IRecognizerParams::LanguageDetectionMode* is now used instead of *IRecognizerParams::DetectLanguage*. *IRecognizerParams::DetectLanguage* is marked as deprecated.

For purposes of compatibility, new property value is changed whenever you change the *DetectLanguage* property: if you change it to TRUE, the property is set to TSPV_Yes, if you change it to FALSE, this property is changed to TSPV_No.

Camera OCR and Color Filtering modules are available for all customers.

CHANGES IN BEHAVIOR

There are no changes in behavior.

NEW FEATURES AND IMPROVEMENTS

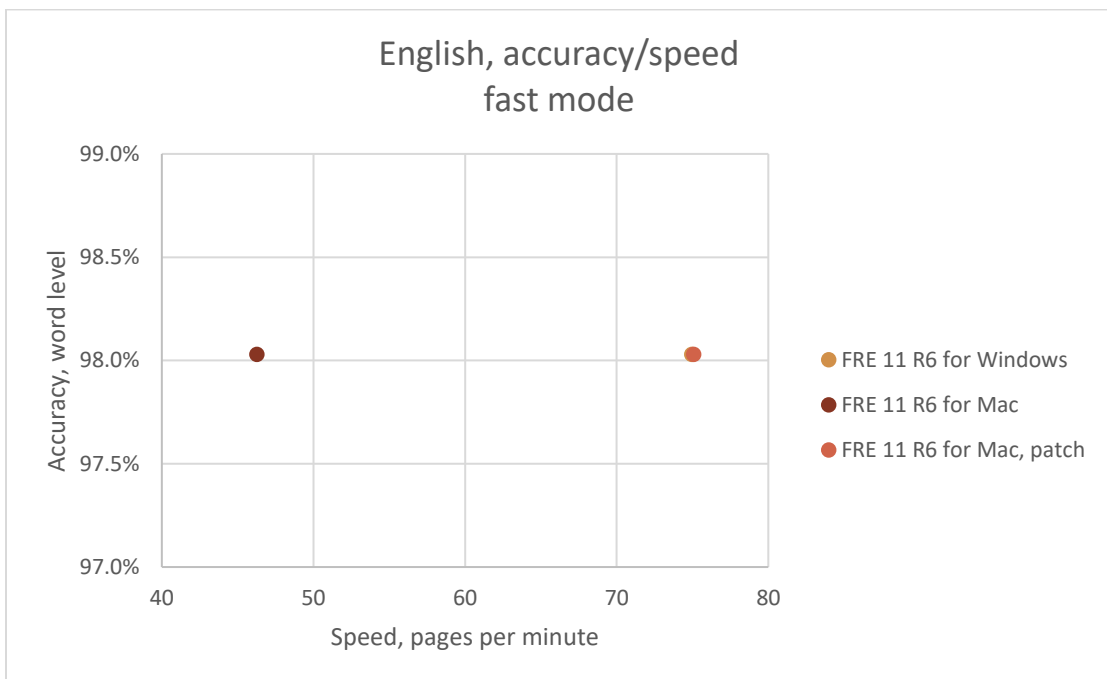
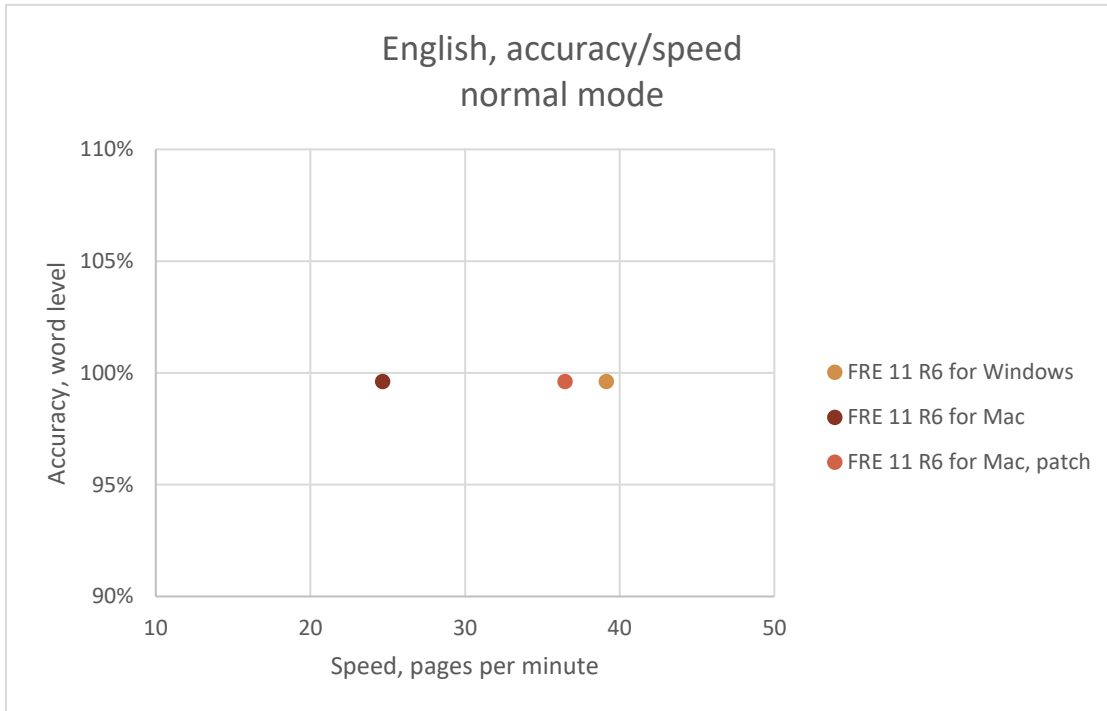
RELEASE 6 UPDATE

There are no new features.

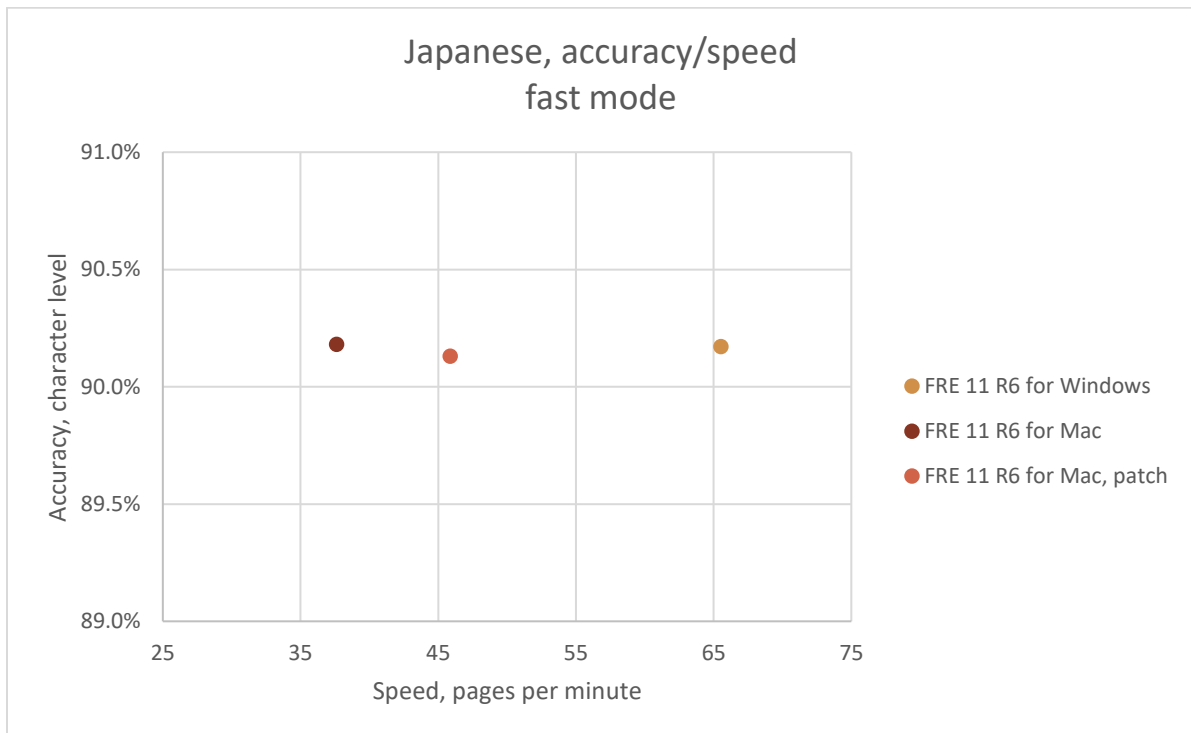
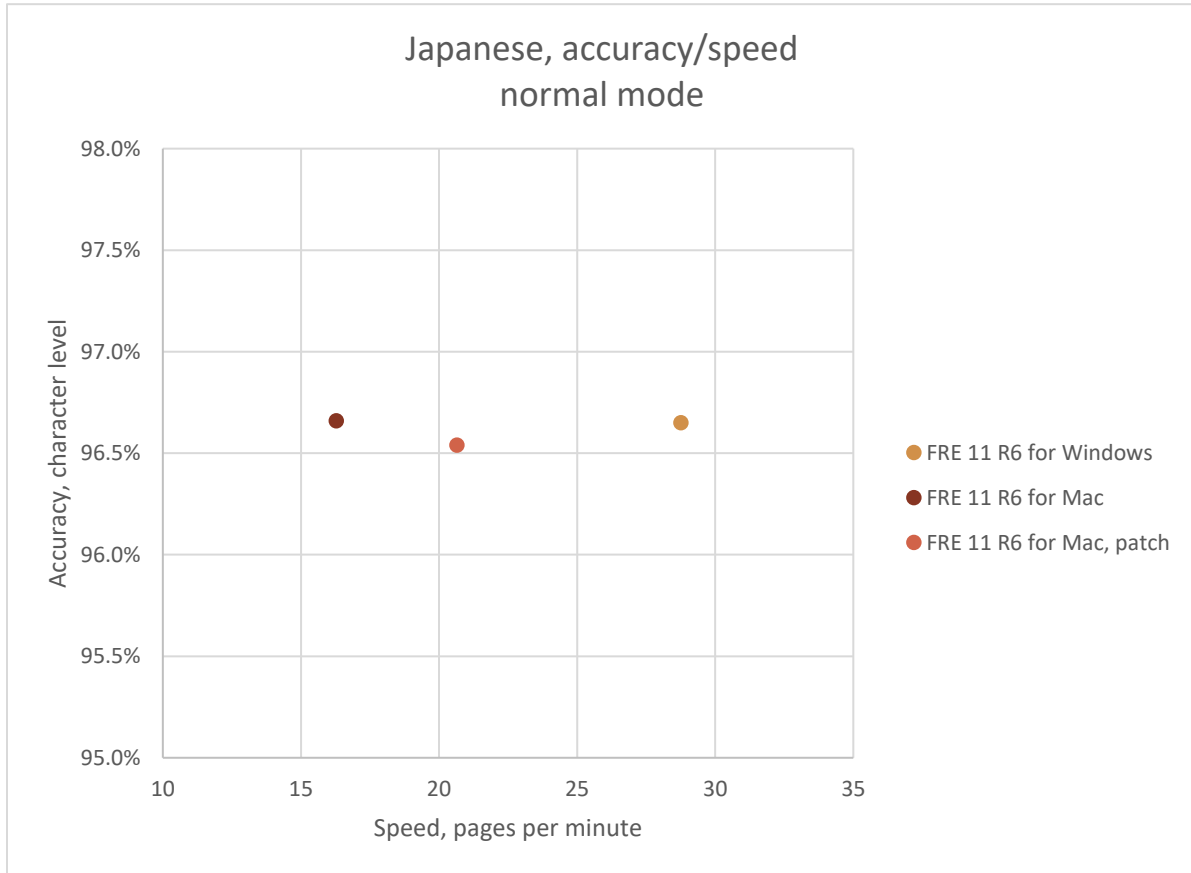
PERFORMANCE RESULTS

This section contains performance results of the release comparing to the previous releases. The processor of the testing machine is Intel® Core™ i5-3450 CPU (3.10GHz, 4 physical cores) with 8GB of RAM.

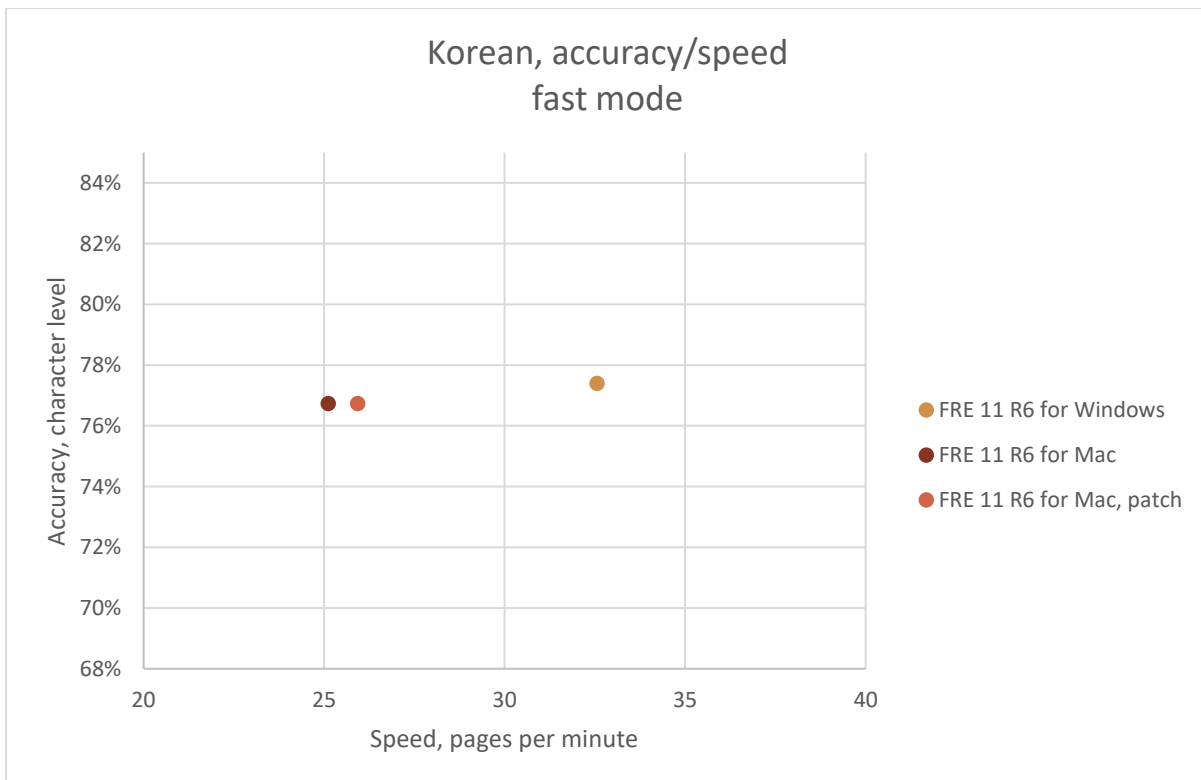
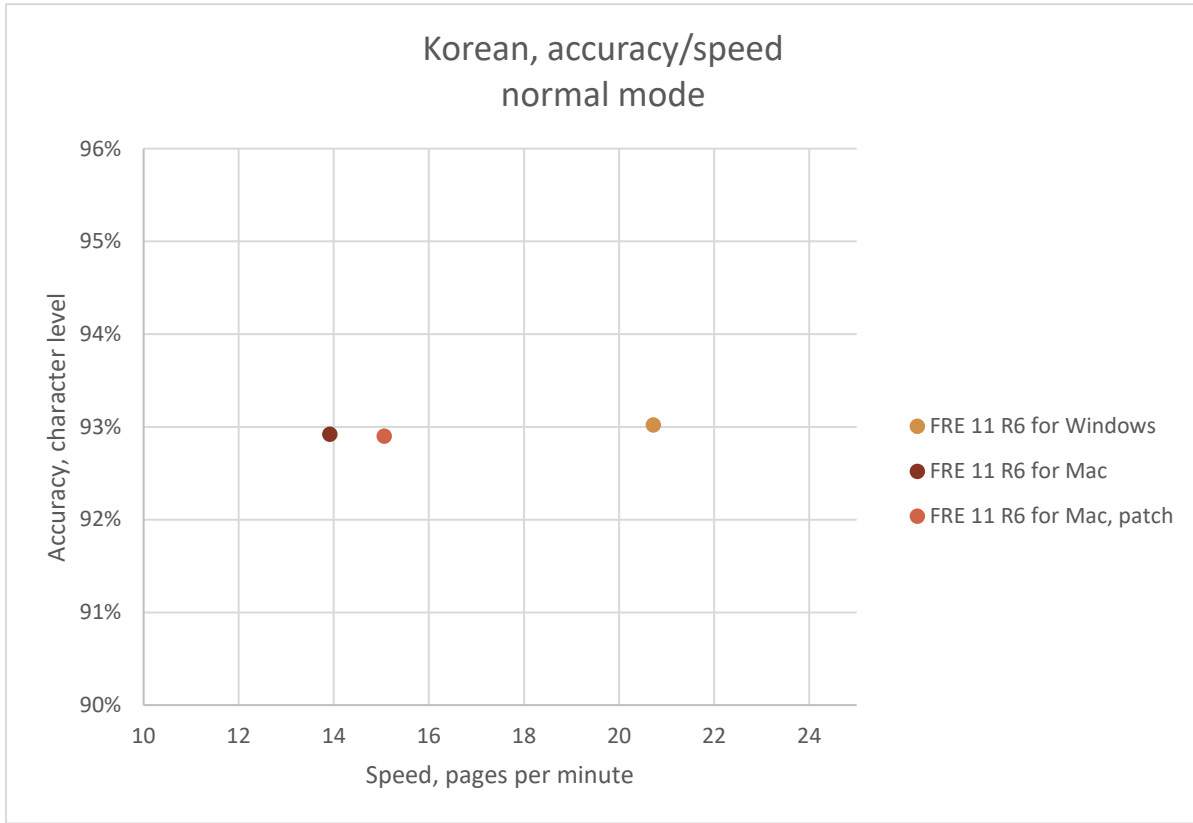
ENGLISH

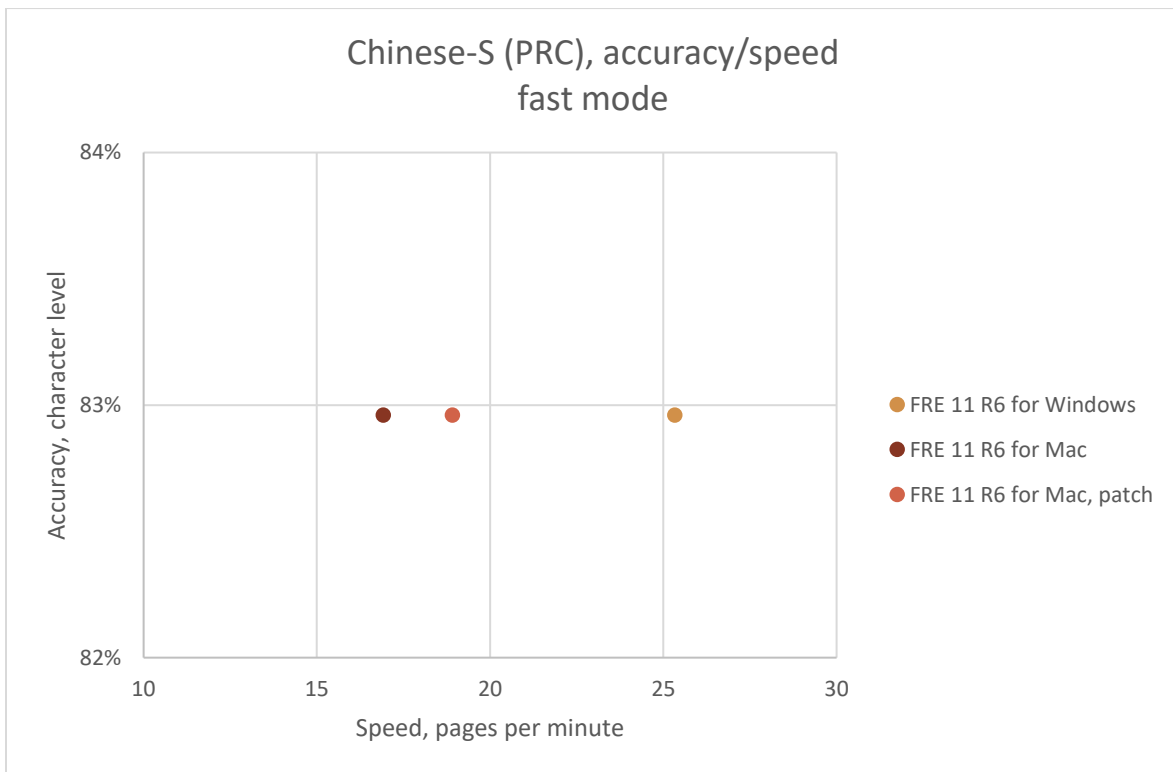
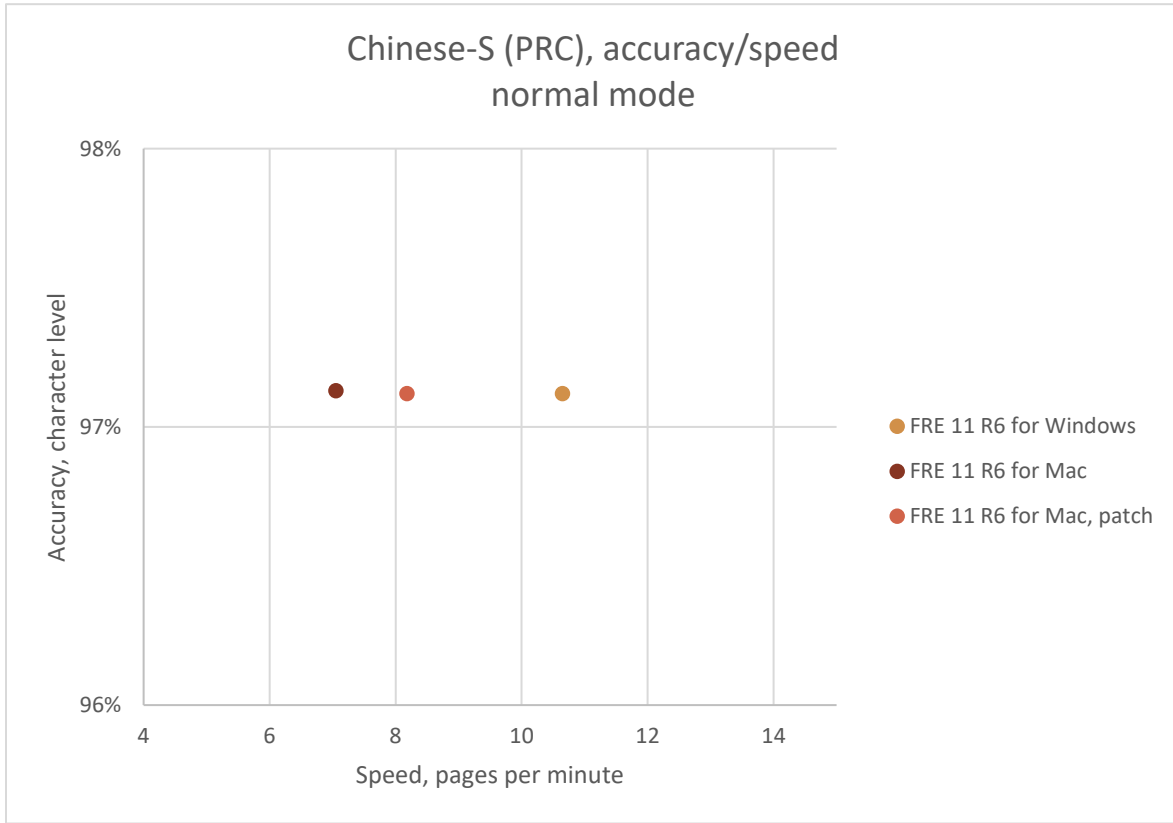


JAPANESE

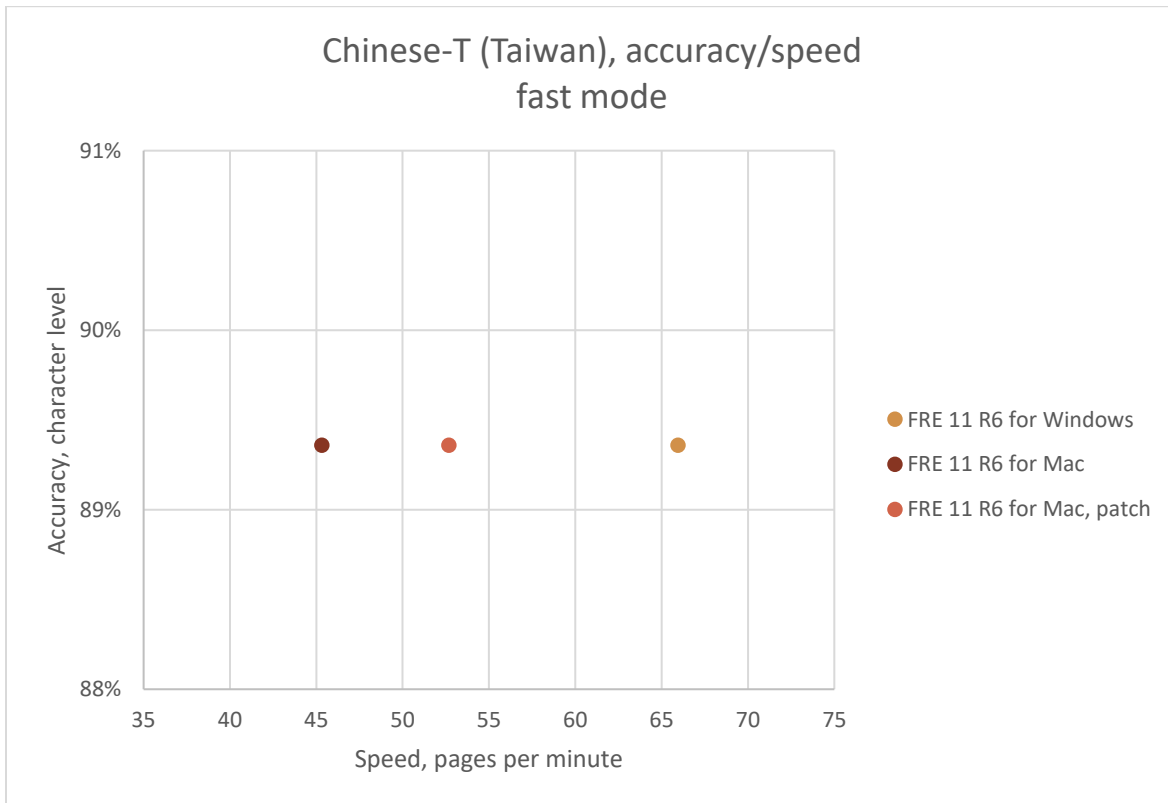
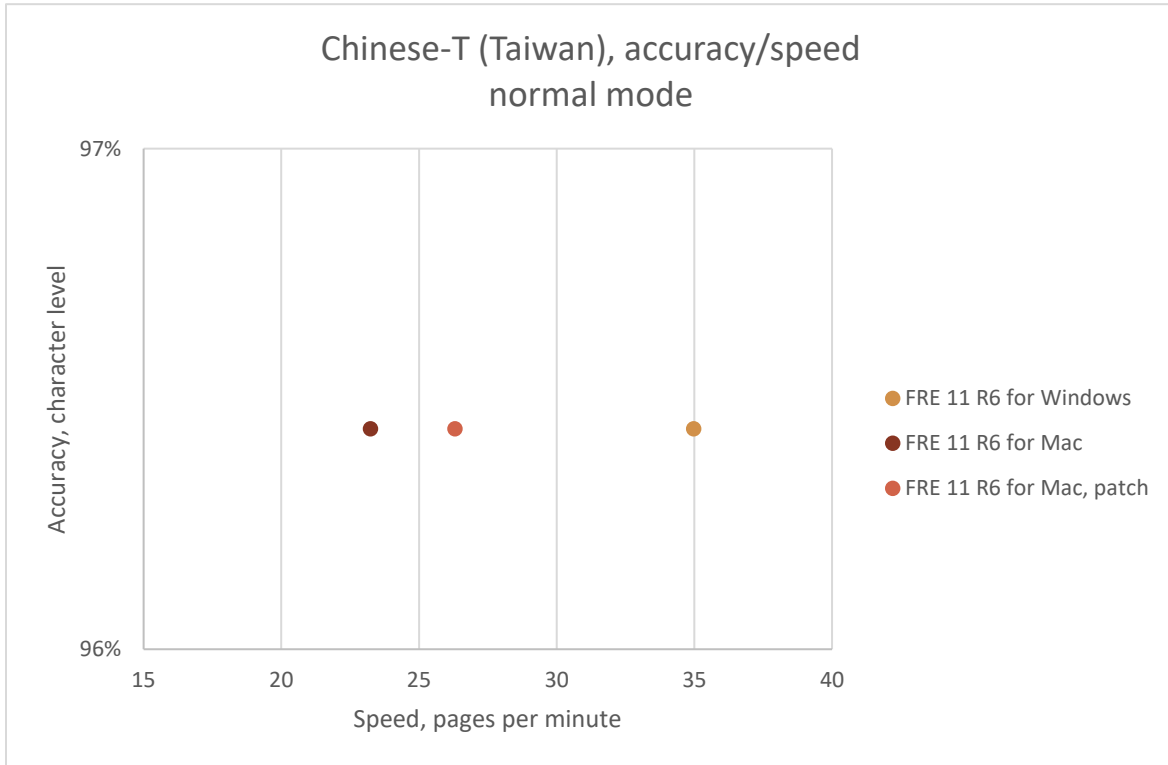


KOREAN

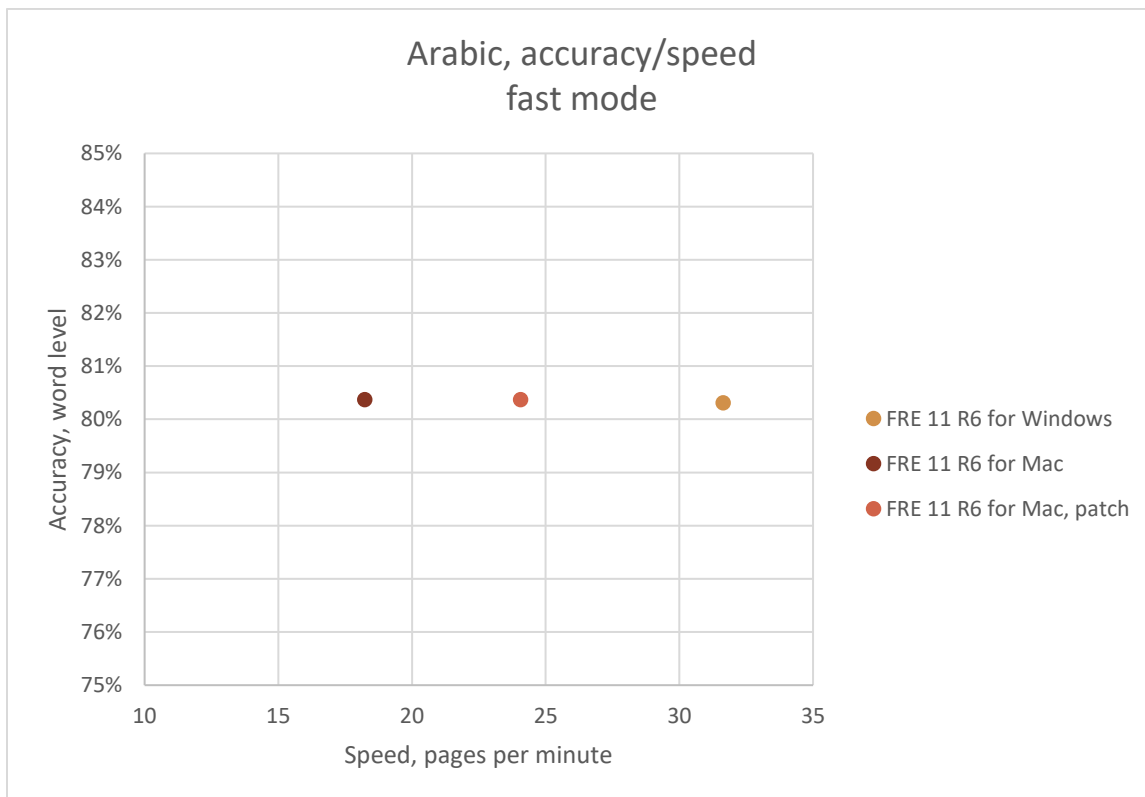
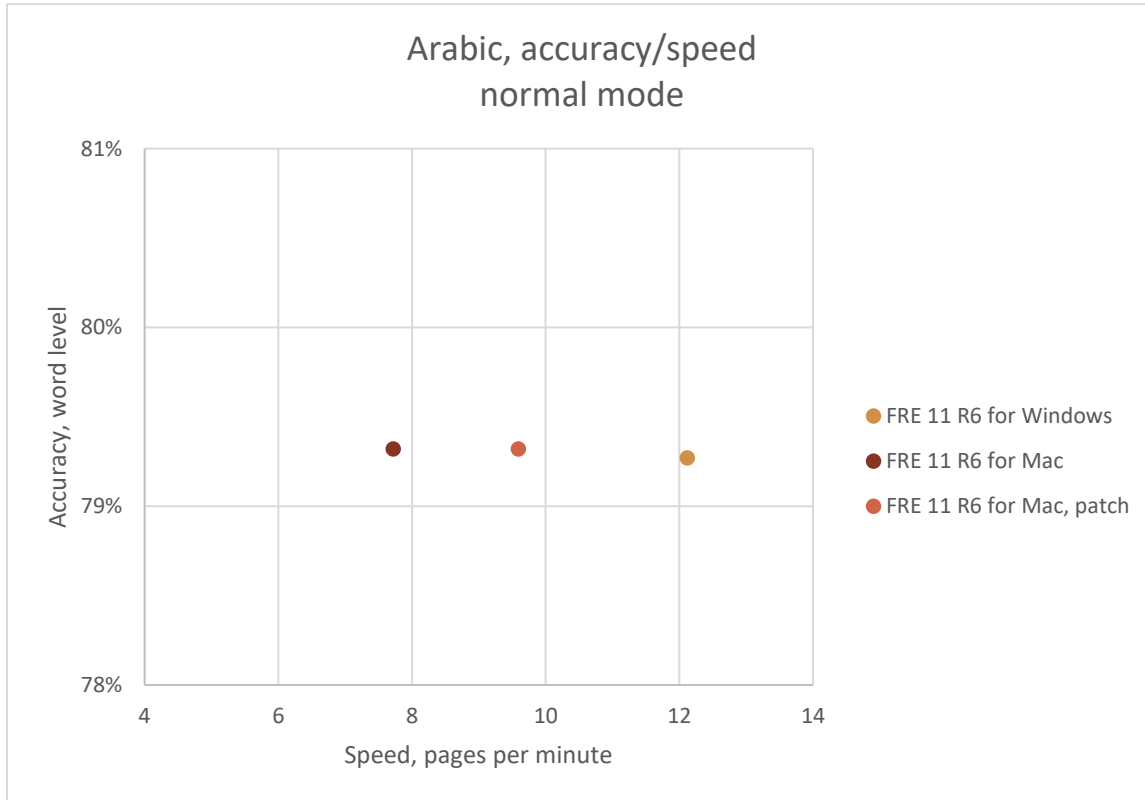




CHINESE TAIWAN



ARABIC



FIXED BUGS

No specific for Mac version bugs were registered and fixed.

KNOWN ISSUES AND WORKAROUNDS

MAC OS VERSION LIMITATIONS

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

- DjVu opening
- Scanning
- ICR/OMR
- Visual Components and other GUI elements
- WDP/WIC/BITMAP input formats and other Windows-specific functionality
- PDF text layer reusing
- Attachments extraction from PDF
- Bookmarks extraction from PDF
- Metadata extraction from PDF (the information about author, keywords, subject and title)

JAVA WRAPPER IS NOT INCLUDED INTO THE DISTRIBUTION

Though we have partly functional Java wrapper for the Engine it still misses some important parts.

In some cases current version of the wrapper is enough, so please consult with HQ product analyst in case of urgent need.

SOME API IS NOT IMPLEMENTED

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

- 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.

FARSI LANGUAGE IS BASED ON ARABIC LANGUAGE

Farsi OCR output indicates Language ID as Arabic (CharParams::LanguageId = LI_ArabicSaudiArabia). This is because Farsi technology preview is based on Arabic OCR language.

The fix will be available in FRE 12.

IPDFMRCPARAMS::MONOCHROMETEXT DOESN'T WORK CORRECTLY

Different algorithms of compressions during export to PDF are used with IPDFMRCPParams::MonochromeText set to FALSE or DEFAULT whereas default value is FALSE.

LANGUAGE AUTO DETECTION USES CJK RESOURCES THOUGH NONE IS SELECTED FOR OCR

'Cjk.*' resource files are used while RecognizerParams::LanguageDetectionMode = TSPV_Yes even though recognition languages set does not include any of CJK languages.

The fix will be available in the FRE 11 R7. The proxy feature description is missing in the documentation

30% SLOWDOWN IN ALL SCENARIOS

Internal tests shows that this release got 30% slowdown comparing to R2.

We are investigating the reason and working on a patch.

ADDIMAGEFILEFROMMEMORY DOES NOT OPEN PDF FILES

An attempt to open a PDF file from a memory ends up with the error 'The image file you specify is empty.'

R7 will include a fix.

R8 GM –

PART NUMBER, BUILD NUMBER

Part#	1161/27
Build#	11.1.19.872047

UPGRADE FROM THE PREVIOUS VERSIONS AND RELEASES

BINARY INCOMPATIBILITY

It is necessary to recompile host application regardless a version of Engine previously used.

API CHANGES

There are no API Changes.

CHANGES IN RELEASE 8 UPDATE

ENGINE::INJECTTEXTLAYER METHOD IS MARKED AS DEPRECATED

IEngine::InjectTextLayer is deprecated and will be removed in the next major release. IEngine::InjectTextLayerEx method must be used instead.

Please find more information on 'Large document conversion to searchable PDFs improvement'.

SOME METHODS, OBJECTS AND INTERFACES RELATED TO ONE-PAGE DOCUMENT PROCESSING ARE MARKED AS DEPRECATED

The following methods are marked as deprecated and will be removed in the next major release:

Methods:

- IEngine::ProcessPage
- IEngine::ProcessPagesEx
- IEngine::ExportPage
- IEngine::ExportPagesEx
- IEngine::SynthesizePagesEx
- IEngine::OpenImageFile
- IEngine::PrepareImageFile
- IEngine::CreateLayout
- Objects:
- DocumentAnalyzer
- DocumentInfo
- Exporter

Interfaces:

- IDocumentAnalyzerEvents
- IRecognizedPages
- IExporterEvents

One-page document processing may be done by using properties and methods of Engine, FRPage and FRDocument objects as following:

One-page API (deprecated)	API to be used	Description
IDocumentAnalyzer interface is a basic element of one-page API. It is used to do the majority of operations with documents.		

PreprocessAnalyzeRecognizePage method	IFRPage::PreprocessAnalyzeRecognize	
PreprocessPage method	IFRPage::Preprocess	
CorrectGeometricalDistortions method	IFRPage::CorrectGeometricalDistortions	
DetectOrientation method	IFRPage::DetectOrientation	
FindPageSplitPosition method	IFRPage::FindPageSplitPosition	
AnalyzePage method	IFRPage::Analyze	
ExtractBarcodes method	IFRPage::ExtractBarcodes	
AnalyzeRegion method	IFRPage::AnalyzeRegion	
AnalyzeTable method	IFRPage::AnalyzeTable	
RecognizePage method	IFRPage::Recognize	
RecognizeBlocks method	IFRPage::RecognizeBlocks	
RecognizeImageAsPlainText method	Recognize using IFRDocument::Process, then call IFRDocument::PlainText attribute	
RecognizeImageDocumentAsPlainText method	Recognize using IFRPage::PreprocessAnalyzeRecognize, then call IFRDocument::PlainText attribute	The main difference from RecognizeImageAsPlainText is that the first method creates a document IImageDocument inside the method, but in the second case the method gets the document as an input parameter.
PreprocessAnalyzeRecognizePagesEx	IFRDocument::Process	IRecognizedPages is

method		deprecated and IFRDocument already have a functionality to manage with multi-page documents
PreprocessPagesEx method	IFRDocument::Preprocess	
AnalyzePagesEx method	IFRDocument::Analyze	
RecognizePagesEx method	IFRDocument::Recognize	
LearnCheckmarks method	IFRPage::LearnCheckmarks	
CleanRecognizerSession method	The method moves to IEngine::CleanRecognizerSession	
AddWordToCacheDictionary method	At this moment there is no analogue for this methods. We are considering to implement them inside the IEngine interface.	
AddWordsToCacheDictionary method		
CleanCacheDictionary method		
AutoCleanRecognizerSession attribute	The attribute moves to IEngine::AutoCleanRecognizerSession	
IDocumentAnalyzerEvents interface is a callback-interface for analysis of events appearing during the documents processing.		
OnRegionProcessed method	IFRPageEvents::OnRegionProcessed	
OnProgress method	IFRPageEvents::OnProgress IFRDocumentEvents::OnProgress	
OnWarning method	IFRPageEvents::OnWarning IFRDocumentEvents::OnWarning	
IDocumentInfo interface is an object that contains an information about the document that was get during the document processing.		

Метод AddImageDocument	IFRDocument::AddImageDocument	Is a wrapper on IFRDocument. The specified IImageDocument document is referred to it and so it is possible to get an information about this document.
Метод DocumentContentInfo	IFRDocument::DocumentContentInfo	
IEngine interface is a mainpoint of entry for work with FREngine		
ProcessPage method	IFRDocument::Process	Methods with *PagesEx suffix use outdated IRecognizedPages
ProcessPagesEx method		
ExportPage method	IExporter::ExportPage	
ExportPagesEx method	IExporter::ExportPageEx	
SynthesizePagesEx method	IFRDocument::Synthesize	
OpenImageFile method	IFRDocument::AddImageFile	
PrepareImageFile method		
CreateLayout method	No implementation	It is assumed that document's layout ILayout can't exist without pages.
IExporter is to save recognized text to different formats.		
ExportPage method	IFRDocument::Export	
ExportPagesEx method		IRecognizedPages is used as a mechanism for multi-page documents processing
IRecognizedPages is a callback-interface for multi-page documents processing in on-page API		
PageIds method	This interface is outdated because a mechanism for multi-page	

Layout method	documents processing has already built-in in IFRDocument.
ImageDocument method	
ReleasePage method	
IExporterEvents is a callback-interface that give an opportunity to monitor the process of document export and its progress.	
ReportPercentage method	Is implemented through the IFRDocumentEvents::OnProgress method

TEXTORIENTATION::ISVERTICALLYMIRRORED PROPERTY IS MARKED AS DEPRECATED

TextOrientation::IsVerticallyMirrored property is marked as deprecated and will be removed in future versions. The cause is that no scenarios when the detection of vertically mirrored orientation is necessary have been found.

ADDED IPDFMRCPARAMS:: USEMULTIPLEMASKS

New property for faster PDF files with MRC was added to FRE. Please, find more information on 'New option for faster printing of PDF using MRC'.

ADDED IPREPAREIMAGEMODE:: RASTERIZEFREETEXT

A new method for rasterizing FreeText annotations was added. Find detailed information on 'An opportunity to rasterize FreeText annotations'.

ADDED INTERFACE AND METHOD FOR TUNING NEW INJECTTEXTLAYEREX2 METHOD

ITextLayerInjectionParams interface and CreateTextLayerInjectionParams method was added. Please, find the detailed information on 'Orientation and skew correction on text injection'.

INSERTTAB METHOD OF THE PARAGRAPH OBJECT WAS ADDED

New method inserts the tabulation symbol into chosen text position.

ITABPOSITIONS::ADDNEW IS MARKED AS DEPRECATED AND ITABPOSITIONS::ADDNEWEX METHOD WAS ADDED

The ITabPositions::AddNew method was not implemented and now it returns E_NOTIMPL code.

To replace this method, ITabPositions::AddNewEx was added.

CHANGES IN BEHAVIOR

AN ABILITY TO SAVE LICENSE COUNTERS VALUE ON LICENSE DEACTIVATION

A new section 'ZeroLevel' was added into License Wizard for FRE 11 Mac. The only one available option in this section is ZeroLevel.

If this option is activated, a license counter value will be saved in ABBYY Registration Server during the license deactivation. On further activation counter will take value from ABBYY Registration Server.

NEW FEATURES AND IMPROVEMENTS

RELEASE 8 UPDATE

CONVERSION DOCUMENTS TO SEARCHABLE PDFS AT FULL THROTTLE REQUIRES NO PAGE NUMBERS SETUP ANY MORE

It is possible to convert documents containing undefined number of pages to searchable PDF using ExportFileWriter. The PagesCount parameter must be set to -1 in this case (this parameter is deprecated and will be removed in future versions).

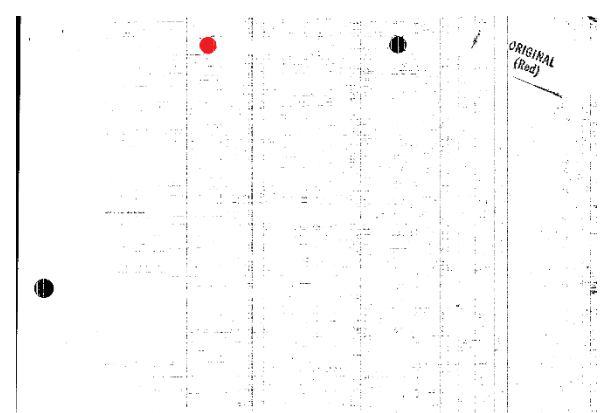
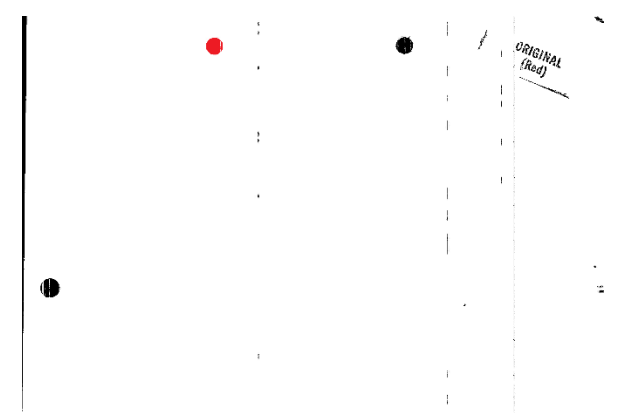
From now, our customers have no necessity to know an amount of pages in the document at the moment of creating a session of recognition.

It can be useful for effective work with scanners when one has to process a lot of pages and doesn't know the amount of them in the document until the end of scanning, but needs the processing to start before the scanning will be finished.

GARBAGE REMOVAL FROM COLOR IMAGES

New method ImageDocument::RemoveGarbageEx works with both color and white-and-black images. From now it is possible to remove garbage from color images using this method.

Input image	Output image
-------------	--------------

<p>ABBYY FineReader Engine 10 Introduction</p> <p>Document Analysis</p> <p>Basic document analysis features</p> <p>Document Analysis is a set of functions for automatic detection of the following objects on a page:</p> <ul style="list-style-type: none"> • Text blocks • Pictures • Tables and table cells • Barcodes • Separators <p>Additionally document analysis provides some special features to prepare image for OCR:</p> <ul style="list-style-type: none"> • process detection of page orientation – 90, 180, and 270 degrees • split double pages • process vertical text detection in table cells 	<p>ABBYY FineReader Engine 10 Introduction</p> <p>Document Analysis</p> <p>Basic document analysis features</p> <p>Document Analysis is a set of functions for automatic detection of the following objects on a page:</p> <ul style="list-style-type: none"> • Text blocks • Pictures • Tables and table cells • Barcodes • Separators <p>Additionally document analysis provides some special features to prepare image for OCR:</p> <ul style="list-style-type: none"> • process detection of page orientation – 90, 180 and 270 degrees • split double pages • process vertical text detection in table cells
<p>упоминалось. Лишь в разделе УП "Законодатель 1, прокурорском надзоре, арбитраже и охране на глава "Судоустройство. Суды Союза ССР". классификаторе отраслей законодательства" в ра дарственного строя" не включена судебная сис 3-м месте, к концу длинного списка отраслей включен п. "Суд. Юстиция. Прокуратура". ни в какой мере не соответствует действитель ту и роли суда в государственном устройстве тигущии Российской Федерации устанавливает,</p>	<p>упоминалось. Лишь в разделе УП "Законодатель 1, прокурорском надзоре, арбитраже и охране на глава "Судоустройство. Суды Союза ССР". классификаторе отраслей законодательства" в ра дарственного строя" не включена судебная сис 3-м месте, к концу длинного списка отраслей включен п. "Суд. Юстиция. Прокуратура". ни в какой мере не соответствует действитель ту и роли суда в государственном устройстве тигущии Российской Федерации устанавливает,</p>
 <p>SECTION 3</p>	 <p>SECTION 3</p>

Dealing with black-and-white images the method works the same way as ImageDocument::RemoveGarbage which will be removed in future versions of FRE.

POSSIBILITY TO INJECT TEXT LAYER TO SELECTED PAGES OF PDF DOCUMENTS

A new method IEngine::InjectTextLayerEx allows to process specific pages in the "image only" or "image on text" PDF documents. It creates a searchable PDF file which contains the same page images and the invisible text layer created from the recognized text of the document.

This new method works the same way as IEngine::InjectTextLayer and has new arguments:

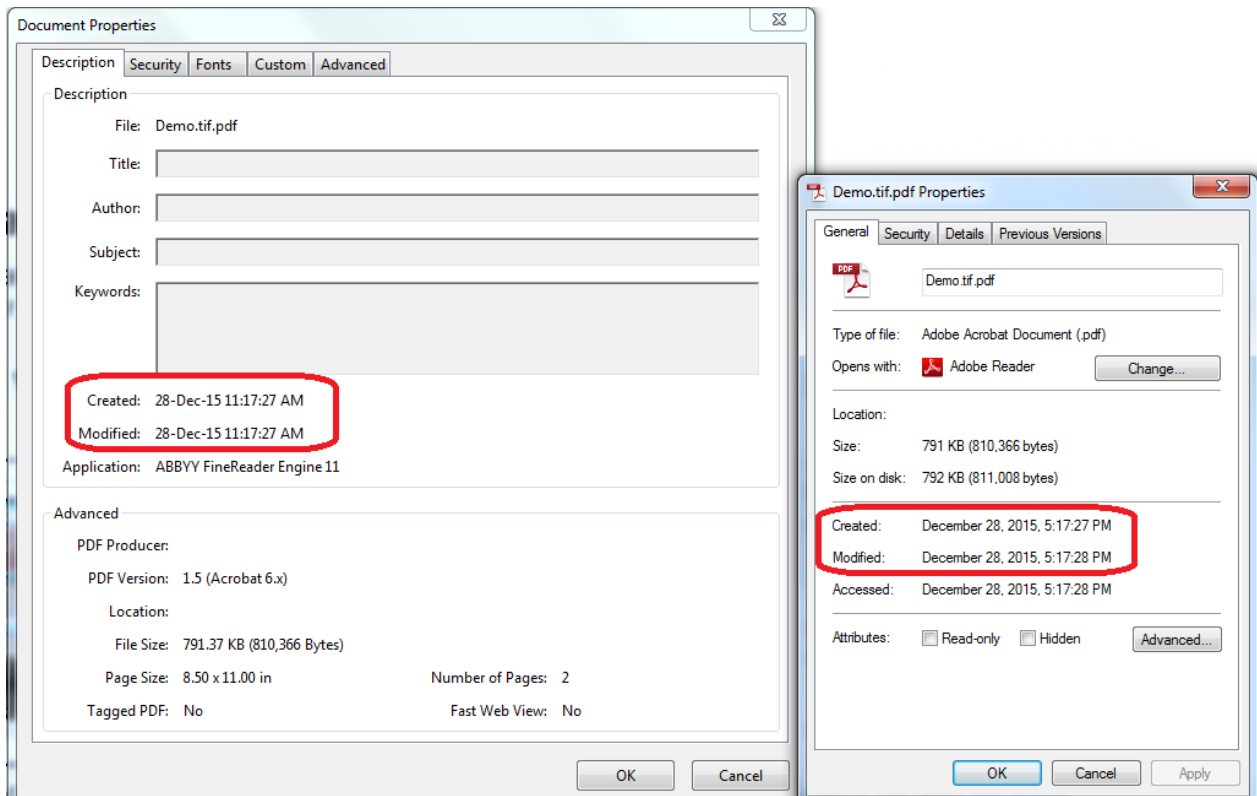
- PageIndices. This parameter refers to the IntsCollection object which specifies the indices of the document pages, to which the text will be injected. This parameter is optional and may be 0, in which case the text will be injected to all the pages of the document.
- ProcessingEvents. Refers to the interface of the user-implemented object that is used for reporting events to the listeners. This parameter may be 0, in which case no callback will be attached.

The IEngine::InjectTextLayer method is now deprecated and will be removed in future versions.

ABILITY TO ADJUST A TIME ZONE FOR PDF EXPORT

New export parameter ReferenceTimeZone of IDPFExportFeatures interface allows our customers to adjust a time zone that will be used for the creation and modification date of the exporting documents.

Earlier FREngine have always wrote modification and creation date using UTC format. It was a problem for one of our customers. Several Adobe products (e.g. Adobe Reader) display creation/modification date of the document without using an information about user's time zone.



To fix this situation the new options were added. ReferenceTimeZone has 3 values: TZT.UTC, TZT.Local and TZT.Daylight values for UTC time, local time and local time considering wintertime respectively. The default value of the parameter is TZT.UTC.

A NEW ARTICLE ABOUT WORKING WITH SCREENSHOTS IN DOCUMENTATION

An article 'I am working with the screenshot image. Are there any special recommendations for screenshot processing?' about suitable settings of FRE for screenshot processing was added. It is in the FAQ section.

ORIENTATION AND SKEW CORRECTION ON TEXT INJECTION

All settings in the PrepareImageMode and PageProcessingParams object such as CorrectSkew and CorrecrOrientation work when passed to InjectTextLayer method.

Sometimes customers have PDFs that contain a mixture of digitally published files and scans. Pages that are digitally published do not need any correction, but scans may need it.

From FRE 11 R8, the new InjectTextLayerEx2 method is available. This method works the same as InjectTextLayerEx, but it additionally can correct orientation and skew of input images.

New TextLayerInjectionParams object allows to tune the parameters of processing the input "image only" or "image on text" PDF files and creation of a searchable PDF file using InjectTextLayerEx2 method.

AN OPPORTUNITY TO CHECK PDF FILES THAT PLACED IN MEMORY FOR TEXT LAYER

Sometimes it is necessary to work with InputStream for working with the files.

Previous versions of FRE 11 has only method IsPdfWithTextualContent for checking PDFs for text layer that required a string for FileName. As the customer works with InputStream it is necessary to pass a byte array as the first parameter for the method. In order to be able to check the files earlier they wrote the stream first in a temp file, but it is not acceptable for them since it would have performance impact.

A new method IsPdfWithTextualContentFromStream accepts reference to a read stream which contains the PDF file in which to detect the text layer.

AN OPPORTUNITY NOT TO WRITE BOM ON EXPORT TO TXT

FRE 11 always wrote Byte Order Mark (BOM) in case of export to TXT.

UTF-8 permits the BOM, but does not require or recommend its use. Byte order has no meaning in UTF-8 and is used only to signal at the start that the text stream is encoded in UTF-8.

Java's UTF-8 encoding has a known behavior and does not recognize this character as a BOM; the result of reading such a stream is a set of characters beginning with FEFF.

So, there are scenarios when clients may need not to write BOM in case of export to TXT. New export option WriteBomCharacter that solves this problem is available since FRE 11 R8.

HUMAN-READABLE NAMES OF PARAGRAPH STYLE ON EXPORT TO XML

Default values of names of paragraph styles were modified and now they can be easier read by a human. Now the names are formed basing on a role of the paragraph and modifications, applied to the style. Earlier GUID was used for naming the styles and it was not informative.

Users also have an opportunity to set paragraph style name manually using `IParagraphStyle::Name` method.

ADDED 'NOT ENOUGH DISK SPACE' ERROR CODE

A new error message 'Not enough disk space' for cases when there is not enough space on the disk was added.

NEW OPTION FOR FASTER PRINTING OF PDF USING MRC

Some our customers complained on too slow printing of PDFs with Mixed Raster Content.

FRE 11 R8 has new option `UseMultipleMasks` of `IPDFMRCParams` for tuning MRC parameters. This option activates a special mode in which different monochrome masks are used instead of one multicolor. This leads to faster document printing.

This mode allows to create only documents with monochrome characters. E.g. it is not possible to create documents with characters with a gradient filling using the new option.

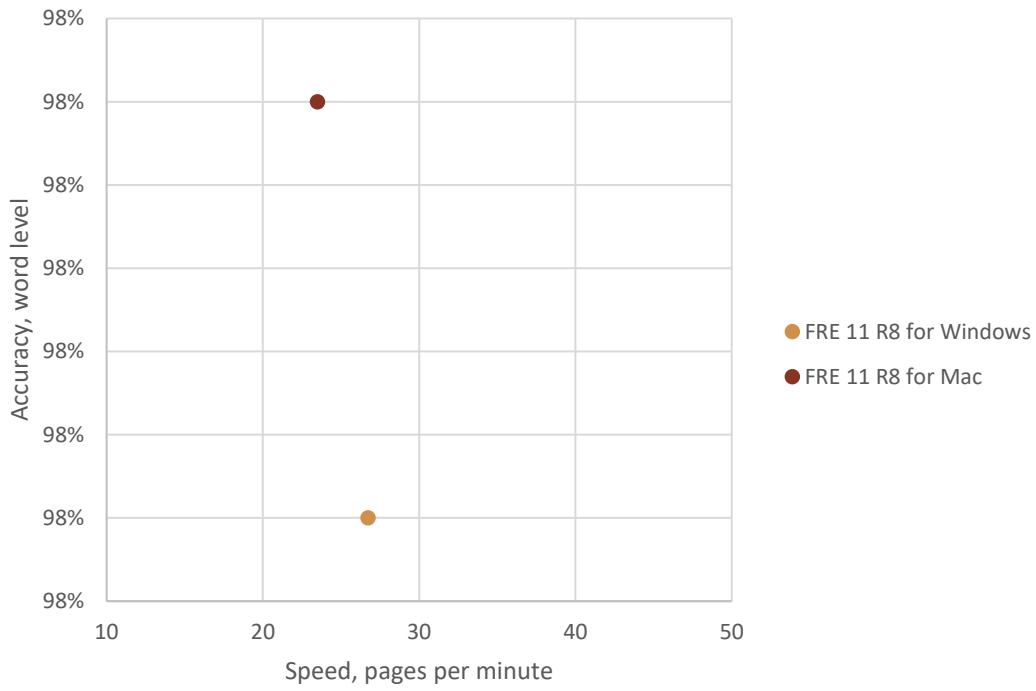
The feature is implemented as technical preview and has not been tested for wide use. So, usage of new mode can cause some issues.

PERFORMANCE RESULTS

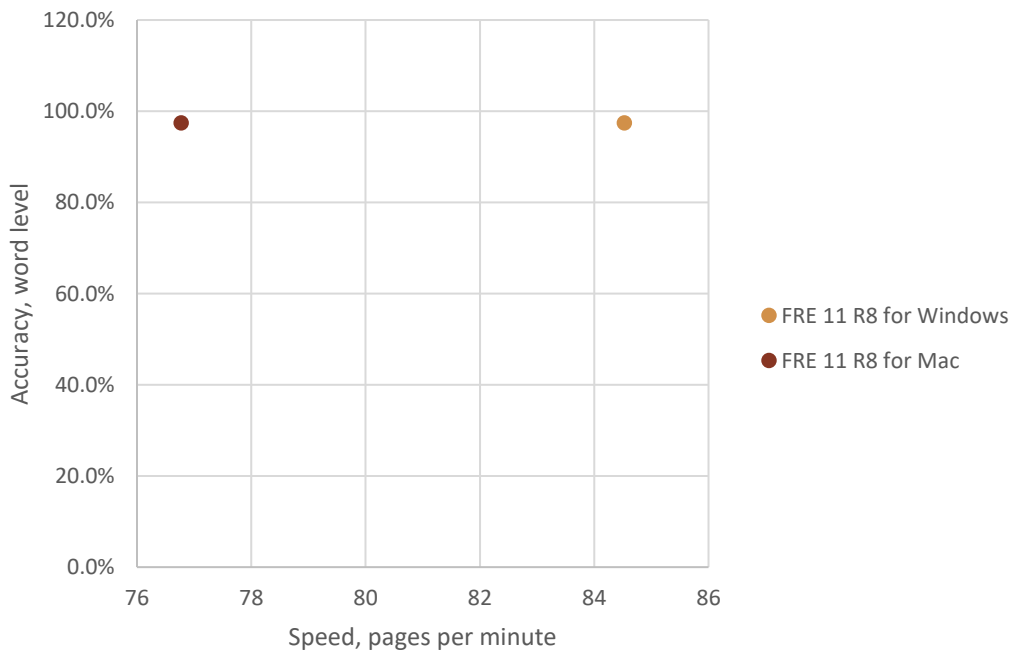
This section contains performance results of the release comparing to the previous releases. The processor of the testing machine is Intel® Core™ i5-3450 CPU (3.10GHz, 4 physical cores) with 8GB of RAM.

ENGLISH

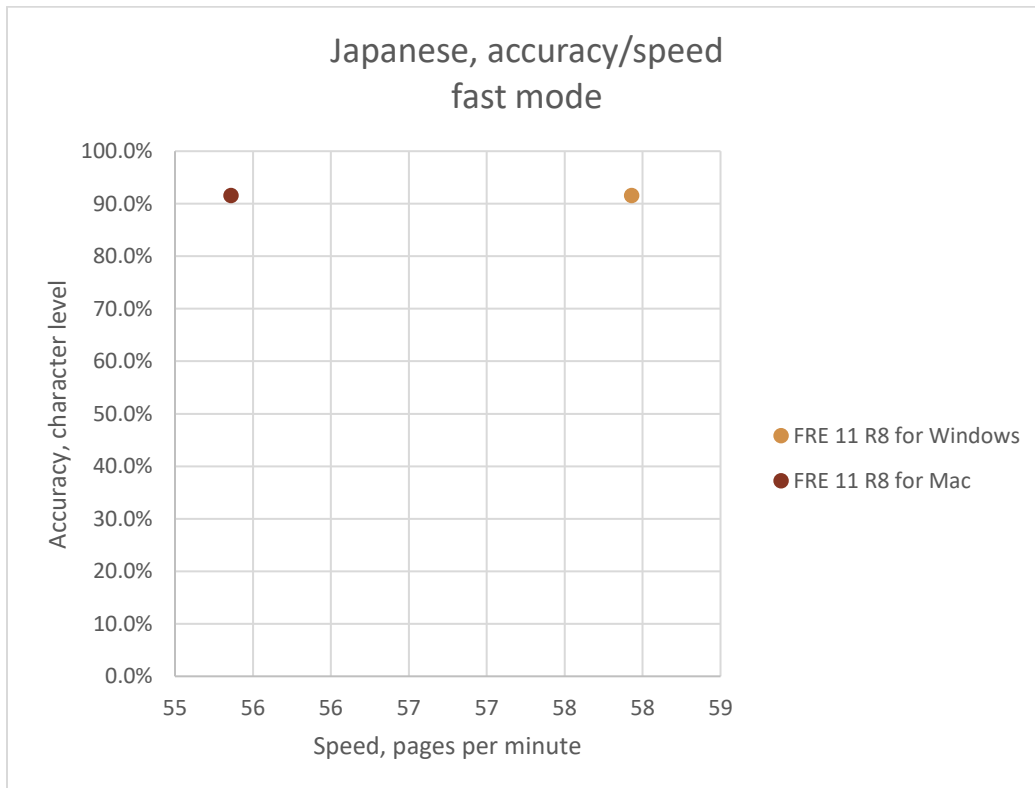
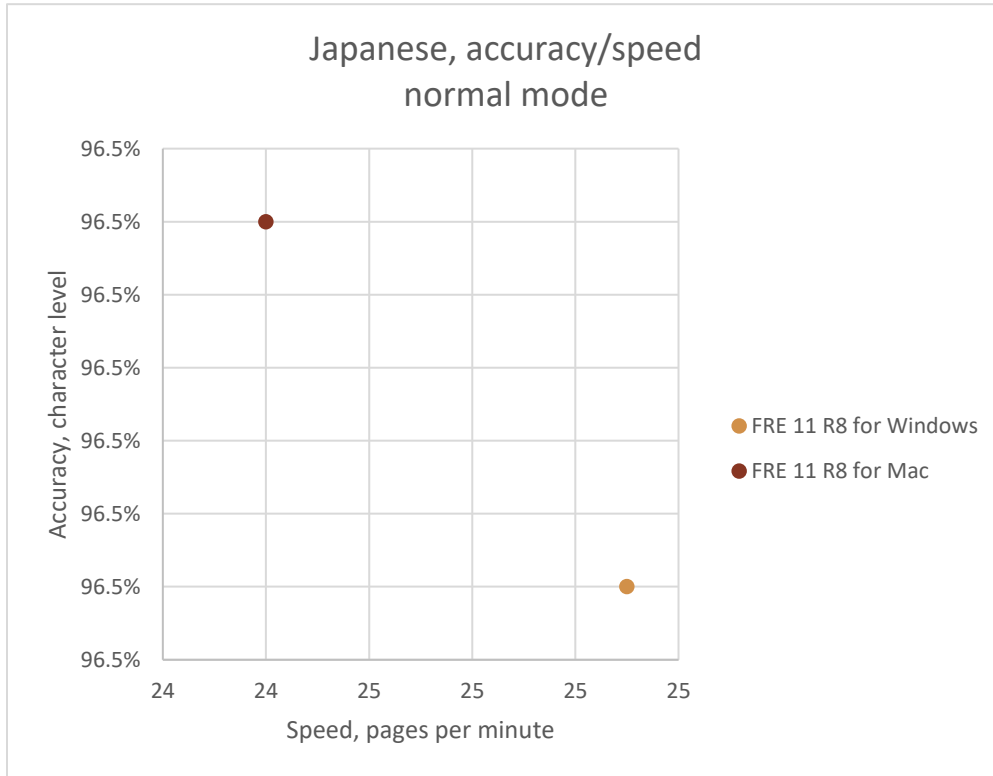
English, accuracy/speed
normal mode



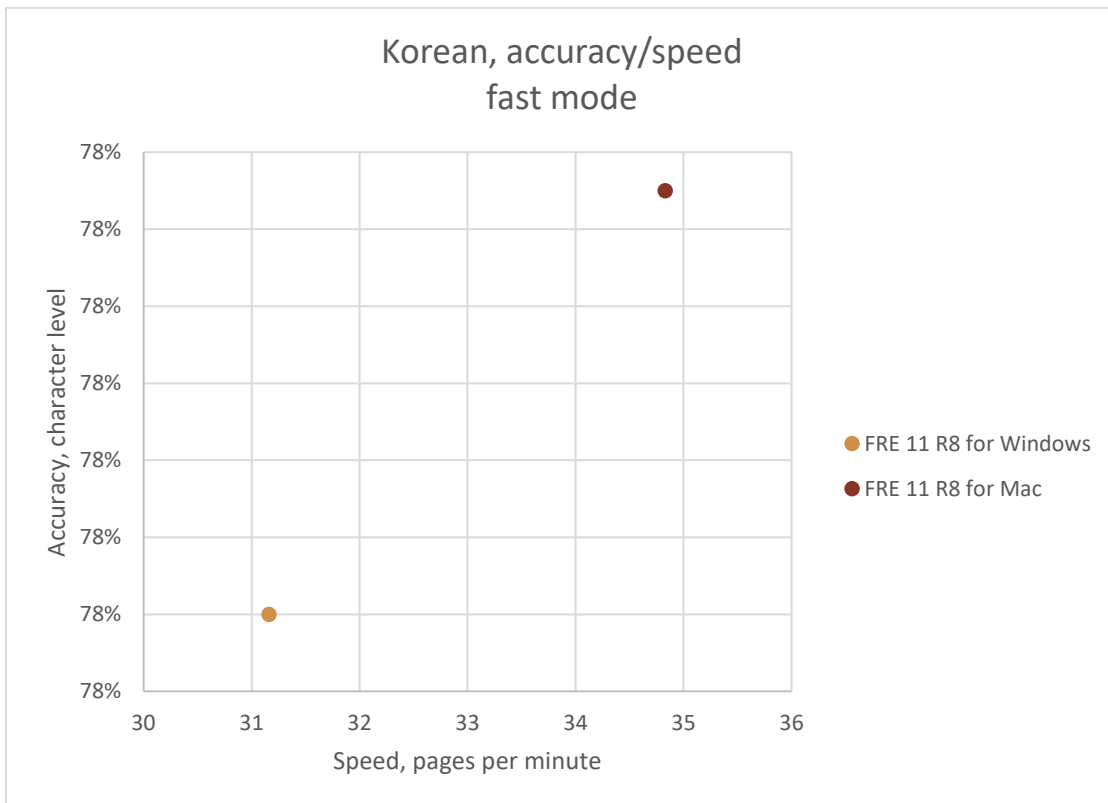
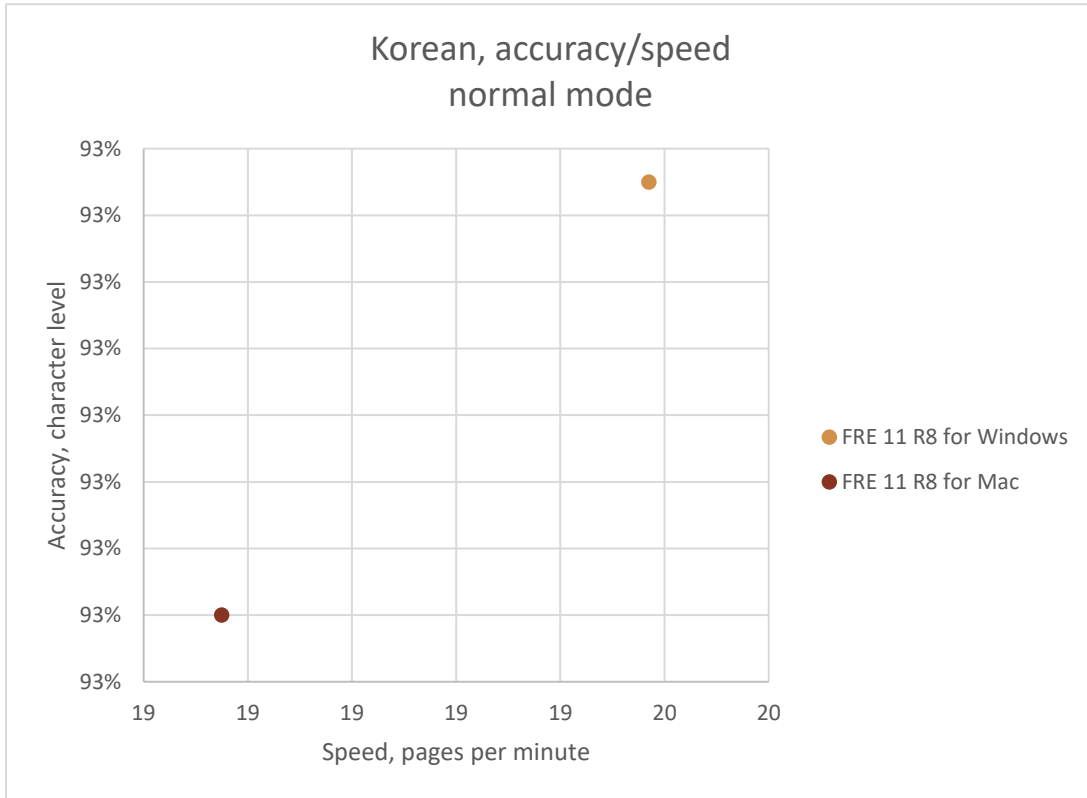
English, accuracy/speed
fast mode



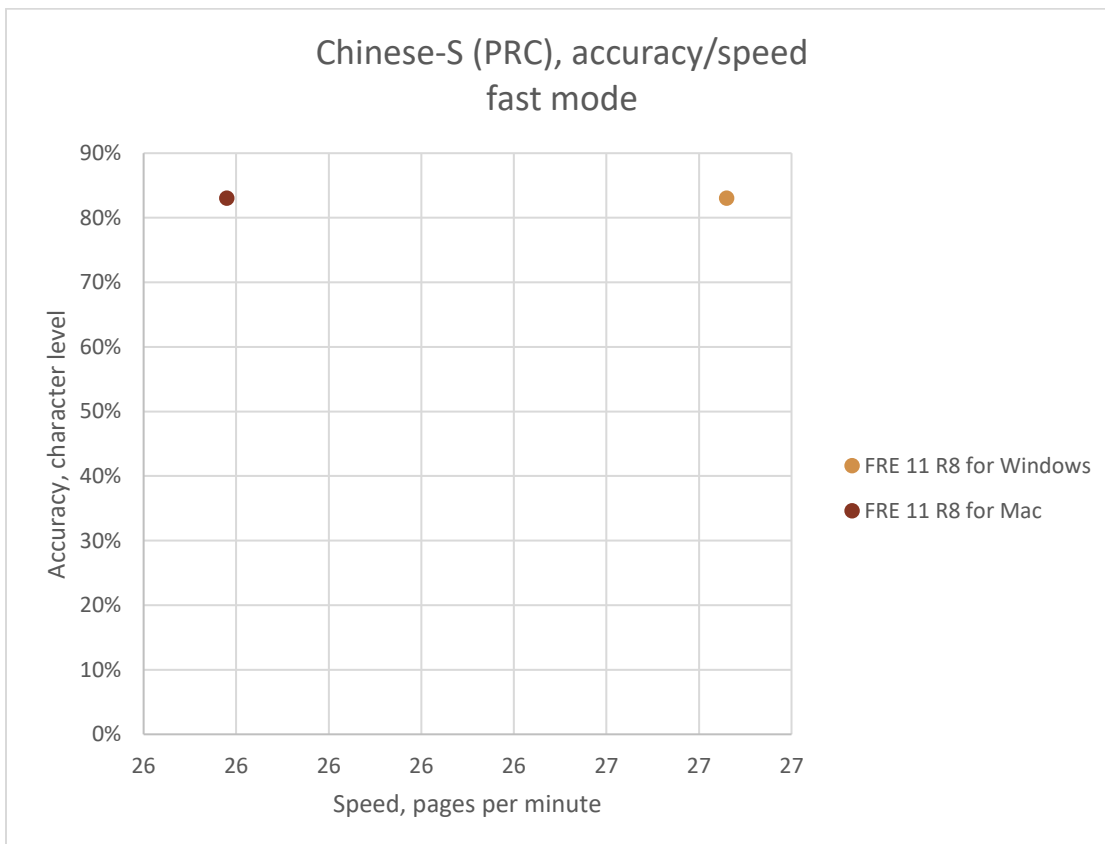
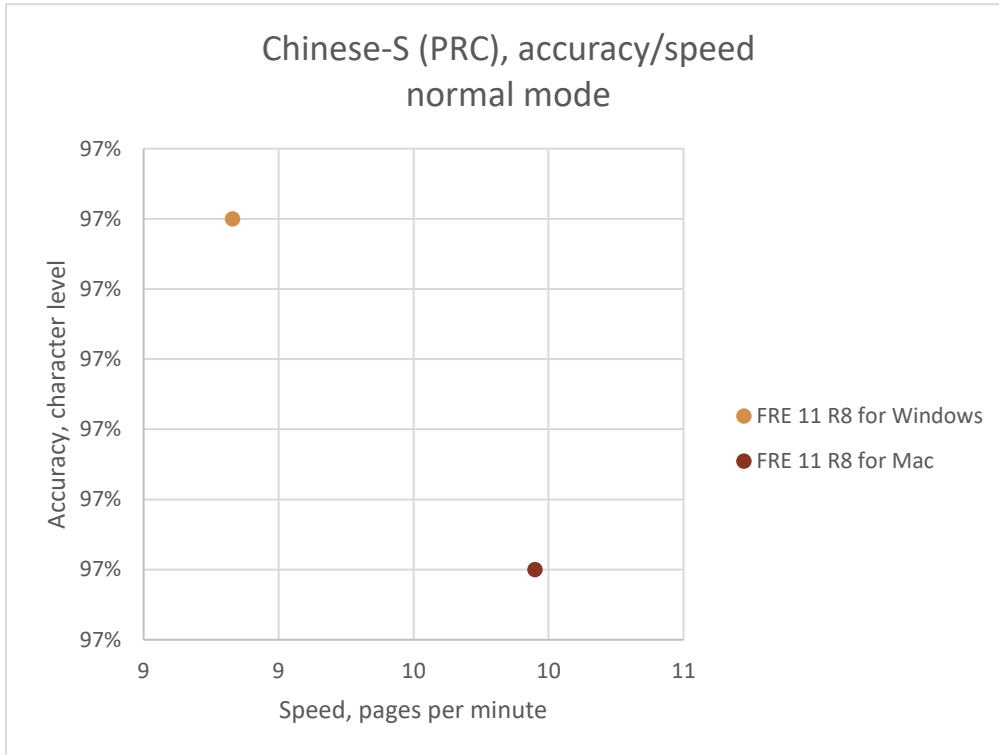
JAPANESE



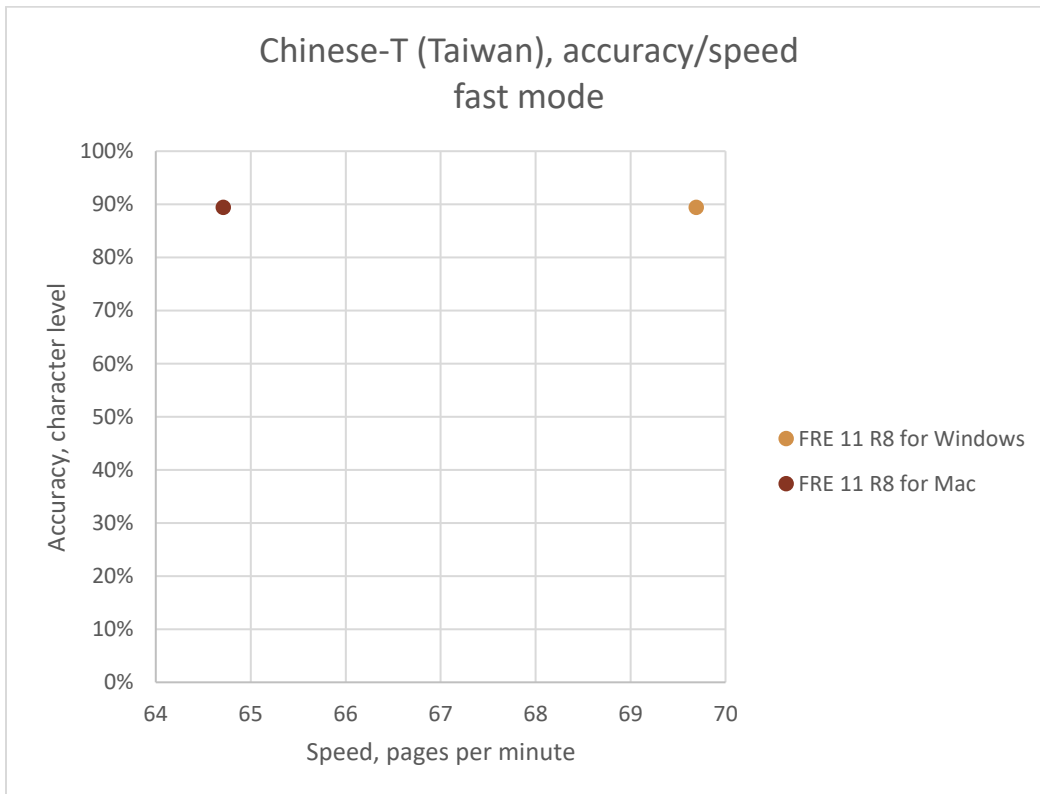
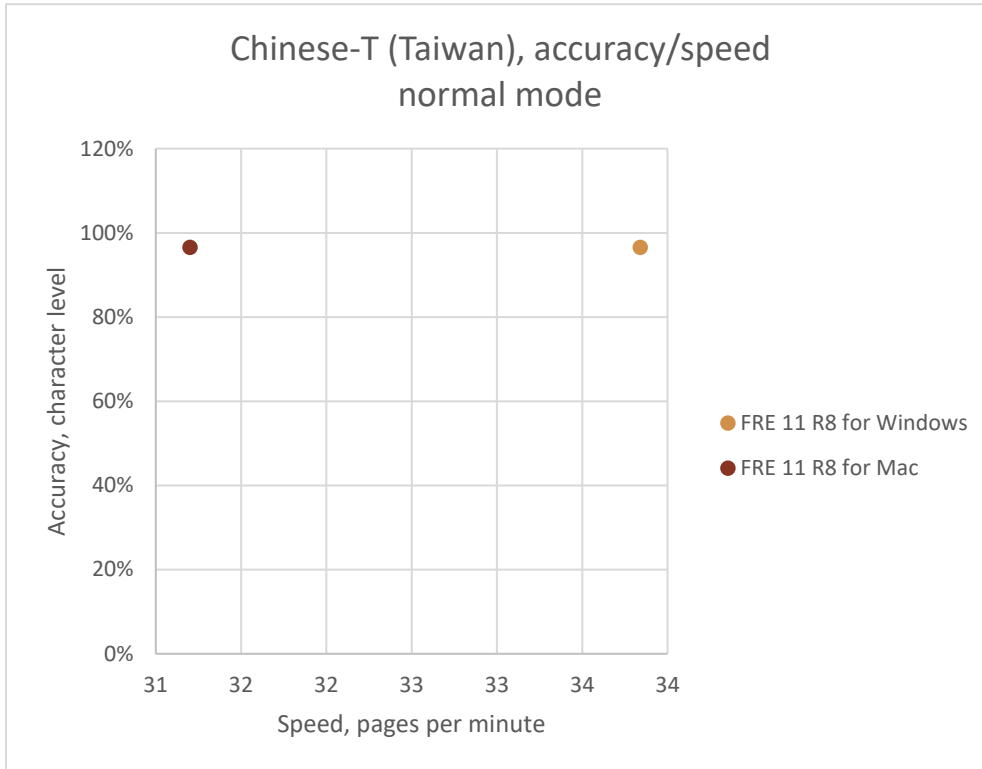
KOREAN



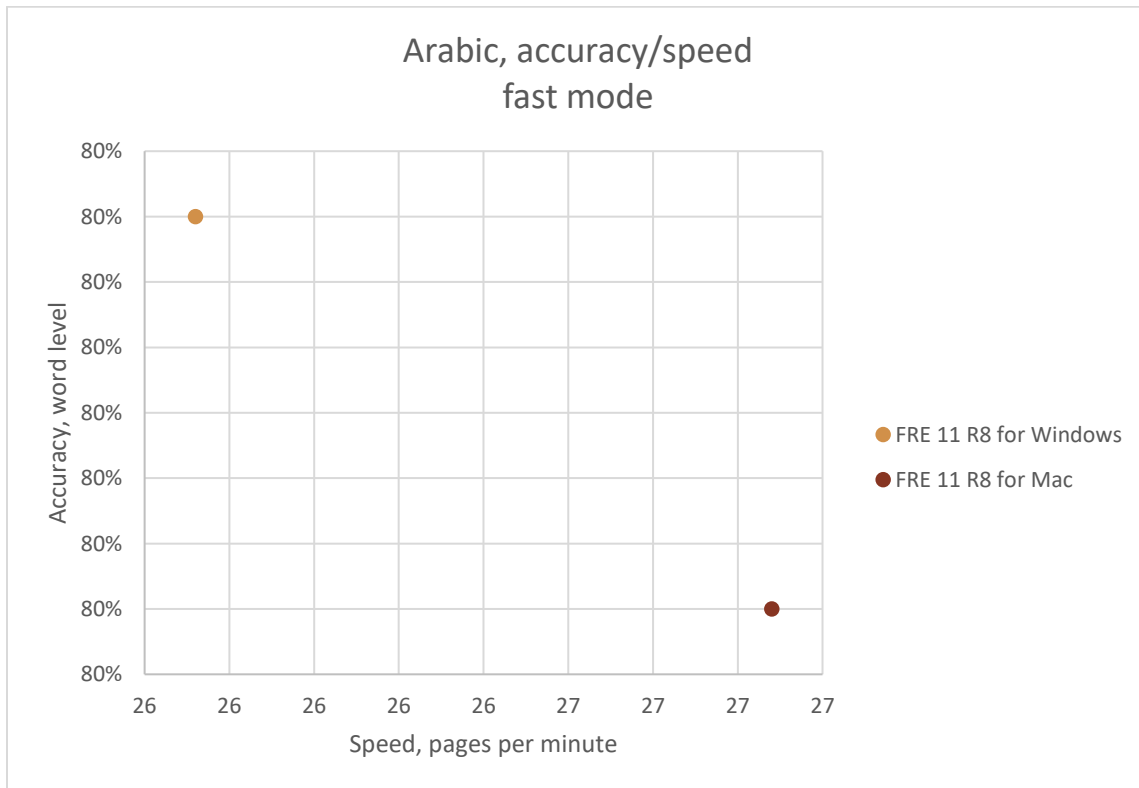
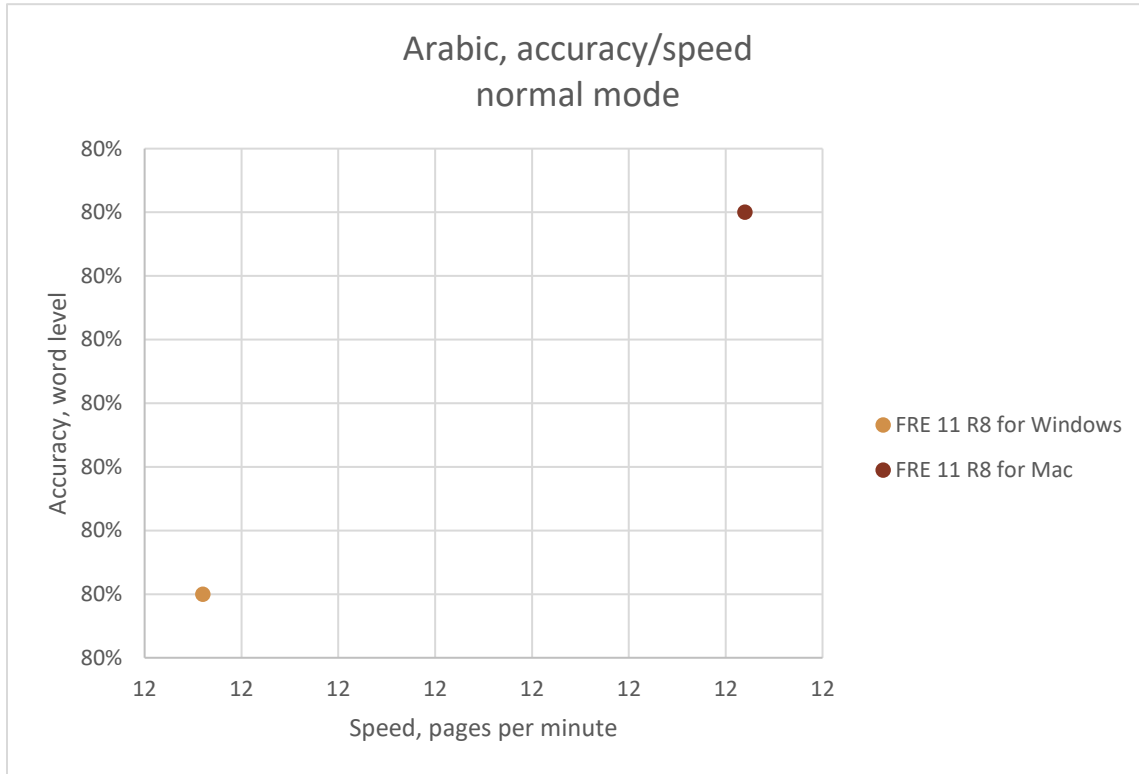
CHINESE PRC



CHINESE TAIWAN



ARABIC



FIXED BUGS

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

Four-point scale will help you to evaluate the severity of each issue, enabling you to make informed decision on how important updates are for your system.

Critical	A bug that causes crashes or hangings of software. Critical bugs can include access violations, internal program errors, stack overflow, out of memory or other exceptions that can lead to program failure.
Major	A bug that does not cause program failure but affects major functionality of a feature or impairs the system's performance. Major bugs can include disparity of the feature functionality to the internal specifications, memory leaks or data corruption.
Minor	A bug that leads to feature malfunctioning or affects minor functionality of the software. Minor bugs can include recognition errors, missing or lost objects, wrong color detection, incorrect document analysis, license counter errors, etc.
Trivial	A cosmetic issue that does not affect the functionality of the product but can cause inconveniences. Trivial bugs can include Help file errors, log errors, incomplete information in error messages, etc.

The following table contains bugs fixed in this release sorted in descending order of severity. If the bugs have workarounds, root causes or side effects, they will be mentioned in the Description section.

Severity	Description	Technology Subsystem	HD #	Office
Critical	Application failure in case of export PDF processing results to PDF, when Engine is loaded using GetEngineObjectEx and FREngineTempFolder is set by default.	API	586844	3A
Critical	An error occurs in case of adding PDF document with password protection.	API	549575	3A

KNOWN ISSUES AND WORKAROUNDS

MAC OS VERSION LIMITATIONS

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

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

- PDF text layer reusing
- Attachments extraction from PDF
- Bookmarks extraction from PDF
- Metadata extraction from PDF (the information about author, keywords, subject and title)

JAVA WRAPPER IS NOT INCLUDED INTO THE DISTRIBUTION

Though we have partly functional Java wrapper for the Engine it still misses some important parts.

In some cases current version of the wrapper is enough, so please consult with HQ product analyst in case of urgent need.

SOME API IS NOT IMPLEMENTED

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

- `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.
- `Ilncut::TextWrapping`. Always returns "TW_Undefined".
- `IRunningTitlesSeriesText::HasSeparator`. Always returns "false".

The implementation is not planned.

FARSI LANGUAGE IS BASED ON ARABIC LANGUAGE

Farsi OCR output indicates Language ID as Arabic (`CharParams::LanguageId = LI_ArabicSaudiArabia`). This is because Farsi technology preview is based on Arabic OCR language.

The fix will be available in FRE 12.

IPDFMRCPARAMS::MONOCHROMETEXT DOESN'T WORK CORRECTLY

Different algorithms of compressions during export to PDF are used with `IPDFMRCPParams::MonochromeText` set to FALSE or DEFAULT whereas default value is FALSE.

LANGUAGE AUTO DETECTION USES CJK RESOURCES THOUGH NONE IS SELECTED FOR OCR

'Cjk.*' resource files are used while `RecognizerParams::LanguageDetectionMode = TSPV_Yes` even though recognition languages set does not include any of CJK languages.

The fix will be available in the FRE 11 R7. The proxy feature description is missing in the documentation

30% SLOWDOWN IN ALL SCENARIOS

Internal tests shows that this release got 30% slowdown comparing to R2.

We are investigating the reason and working on a patch.

ADDIMAGEFILEFROMMEMORY DOES NOT OPEN PDF FILES

An attempt to open a PDF file from a memory ends up with the error 'The image file you specify is empty.'

R7 will include a fix.