

ABBYY FineReader Engine 12 for Mac

Release 5

Release Notes

---

# Table of Contents

<b>Technical information.....</b>	<b>4</b>
<b>New features .....</b>	<b>4</b>
Technologies .....	4
OCR technology update .....	4
Machine learning barcode recognition technology.....	4
Improved table structure analysis .....	4
New 'Accurate' recognition mode .....	4
OCR improvements for text near stamps and signatures .....	5
Better PDF file processing .....	5
NeoML (open-source ML C++ library) usage.....	5
Embedded PDFium for processing PDFs.....	6
Bangla OCR: New supported language.....	6
Arabic OCR .....	6
Japanese OCR: support of single '°C' unicode character.....	6
PowerPoint export improvements.....	6
Enhanced extraction of PDF attachments .....	6
Using grace periods for licenses.....	6
System requirements .....	6
For supported systems .....	6
Deprecated systems.....	6
New features .....	7
New option for licenses with expiration date.....	7
New library to work with PDFs.....	7
Export to XLS/XPS .....	7
API Changes .....	7
Changes in recognition mode .....	7
Engine API.....	8
Barcode API .....	8
Document-related API .....	8
Image-related API.....	8
Text-related API.....	9

BCR synthesis .....	9
Parameter Objects.....	9
License-related API .....	9
Deprecated and deleted functionality .....	9
Known issues .....	10
Documentation .....	10
Online help .....	10
Corrected e-mail for activating the license.....	10
Better description of Japanese Modern in Recognizing CJK article .....	10
Glossary .....	10
Chinese documentation.....	11
Unsupported PDF formats.....	11

## Technical information

Release	Part #	Build #	OCRT build #	Release date
Release 5 Update 1 for macOS Intel	1377/15	12.5.12.26024	16.2.503.564845	2022.04.25
Release 5 Update 1 for macOS ARM	1377/16	12.5.12.26025	16.2.503.564844	2022.04.25

## New features

### Technologies

#### OCR technology update

- To implement the neural network approaches in OCR technologies, ABBYY FineReader Engine was enhanced by the new features of processing the Latin symbols:
  - New language model using both for a consistent choice of word variants generated by OCR and for the substitution of new word variants.
  - End-to-end recognition for Latin-based languages.

#### Machine learning barcode recognition technology

- Neural network architecture introduces a new barcode recognition model based on parsing image pixels into two categories: barcode or non-barcode. Regions built around the connected components are further considered as a barcode hypothesis. The recognition process is started using data from each region containing the type of barcode that is highlighted as the most probable.
- New technology depends on image resolution and has no dependency on the number of barcode types to detect. This leads to slower processing speed comparing to the legacy technology on images of less than 300 dpi resolution and when a single or few barcode types are to be detected.

#### Improved table structure analysis

- With the improved mechanism of document conversion, ABBYY FineReader Engine can detect false vertical separators and correctly process the tables with columns of numbers in the **‘Accounting’** format when a currency symbol ('\$') is aligned to the left in all cells.

#### New **‘Accurate’** recognition mode

- The new **‘Accurate’** mode allows you to get the maximum quality of the output document, assuming a reasonable slowdown in the recognition speed. To obtain the best results in processing, use this mode on poor-quality documents and images:
  - invoices and contracts (scans, small text, photographs)
  - receipts (poor print quality of the original receipt, scan, or photo)
  - ID documents (text is printed on a complex background with textures or illustrations, holography on the text)

## OCR improvements for text near stamps and signatures

- To improve the results of recognition for the agreements, a new neural network model for detecting stamps, logos, and signatures is now applied. This model allows to detect the additional elements in the footers and requisites area of the document, exclude them from the analysis, and highlight the text in the image, ignoring the details. The recognized blocks are superimposed on the image in such a way as if it were a similar document without extraneous marks and stamps.

Document objects detection (stamps, signatures, logos).  
Examples (DA result with object detection)

ABBYY

- Correct text order, further natural language text analysis is possible
- All text is inside text blocks, can be recognized



19

## Better PDF file processing

- ABBYY FineReader Engine can recognize various PDFs from image-only (scans) to digitally-born (text and pictures) and tries to reuse all information from them. But there are occasions when PDF files are a mix of image-only and digitally-born pages in one PDF file, thus making them difficult to process. To cover such cases and let customers recognize such files in manual mode, ABBYY FineReader Engine provides new options:
  - Adaptive recognition to improve and speed up PDF processing (default PDF recognition mode: **PullTextAndRecognizeRest**).
  - Text layer quality classifier for preserving good one in the output format (the CheckTextLayer method of the IFRDocument object)
  - New content reuse mode for processing the PDF or Office documents with mixed content (**CRM\_ContentAndPictures** in **SourceContentReuseModeEnum**).
- Additionally, ABBYY FineReader Engine was enhanced by the option detecting the presence of a digital signature on a document page or inside the PDF (IFRDocument::SourceHasDigitalSignature, IFRPage::SourceHasDigitalSignature).
- To ensure more flexible forming the PDF contents, ABBYY FineReader Engine offers the new options:
  - Opening PDF Portfolios and processing their contents
  - Adding custom images to the output PDF and managing their positions

## NeoML (open-source ML C++ library) usage

- NeoML is an end-to-end machine learning framework that allows you to build, train, and deploy ML models. This framework is used by ABBYY engineers for computer vision and natural language processing tasks, including image preprocessing, classification, document layout analysis, OCR, and data extraction from structured and unstructured documents.
- Key features:
  - Neural networks with support for over 100 layer types
  - Traditional machine learning: 20+ algorithms (classification, regression, clustering, etc.)
  - CPU and GPU support, fast inference
  - ONNX support
  - Languages: C++, Java, Objective-C

- Cross-platform: the same code can be run on Windows, Linux, macOS, iOS, and Android

## **Embedded PDFium for processing PDFs**

- PDFium is a cross-platform native library conforming to PDF standards and controlling all operations related to PDF, including processing, parsing, rendering, and obtaining the output.

## **Bangla OCR: New supported language**

- Bangla language has been added to ABBYY FineReader Engine as a technical preview and now can be applied in scenarios with simple document layout preservation.

## **Arabic OCR**

- ABBYY FineReader Engine got the new neural network technologies for significantly increasing the Arabic recognition accuracy and correcting the misrecognition of European insertions into Arabic text, losing text strings, and reducing the excessive calls.

## **Japanese OCR: support of single '°C' unicode character**

- To improve the detection of the symbols in Japanese, the output document is corrected by the OCR technologies, so the single '°C' Unicode character goes out instead of the two separate characters '°'.

## **PowerPoint export improvements**

- ABBYY FineReader Engine now has a better conversion for the presentation formats including the enhanced layout preservation and the generating of the correct appearance for the output:
  - Frames preservation around text elements, paragraphs
  - Separators preservation
  - Logical grouping of text elements (e.g., text columns)
  - Text fitting into a shape
  - Preservation of indentation between paragraphs

## **Enhanced extraction of PDF attachments**

- Now, you can obtain the list of attachments from PDF Portfolio in the order established when creating a PDF file. Use the IFRDocument::PDFAttachments property to access the collection of the documents extracted from the input PDF files.

## **Using grace periods for licenses**

- With the new option, customers can use the ABBYY FineReader Engine license for some time after the expiration date, thereby enlarging the license validity period.

# **System requirements**

## **For supported systems**

- macOS 10.15.x Catalina
- macOS 11.x Big Sur

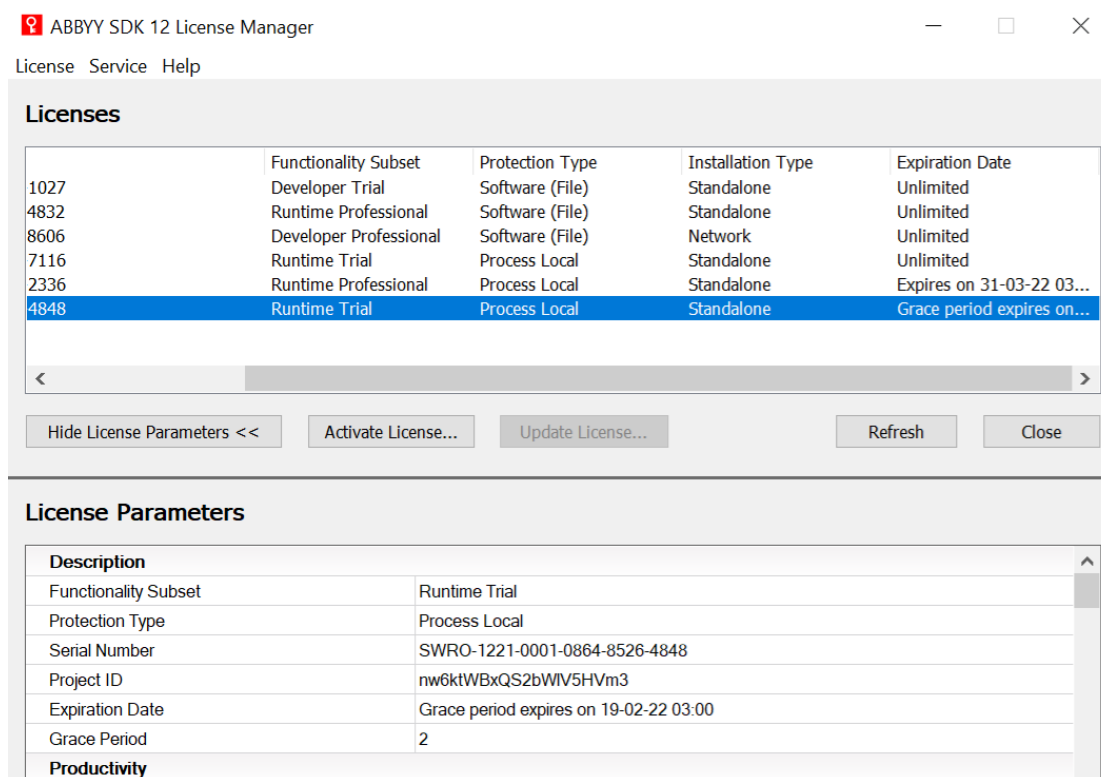
## **Deprecated systems**

- macOS 10.13.x High Sierra
- macOS 10.14.x Mojave

## New features

### New option for licenses with expiration date

- If a license is valid and the expiration date hasn't reached yet, the '**Expiration Date**' field in License Manager contains the exact date value (Expires on <your\_license\_expiration\_date>).



- After passing the expiration date, the grace period starts, during which the license is still valid. The duration of this period depends on a particular license (it displays in the '**Expiration Date**' field for license and in License Parameters). Upon the expiration of the grace period, the license becomes permanently invalid.

### New library to work with PDFs

- Now, all processes connected to PDF processing go through the Google PDFium Library.

### Export to XLS/XPS

- The API required for document export to XLS/XPS format (**XLSExportParams**, **XPSExportParams**, **FileExportFormatEnum** (**FEF\_XLS**, **FEF\_XPS**), **IXLSExportParams::XLFileFormat**, **XLFileFormatEnum**, **XPSExportModeEnum**) is now fully available in Linux version.

## API Changes

### Changes in recognition mode

- **IRecognizerParams** object now has the new '**Mode**' property instead of the '**FastMode**' and '**BalancedMode**' properties. The '**Mode**' property allows to specify the recognition mode to be applied to the target documents - Fast, Normal, or Accurate mode, each of which provides its own accuracy and recognition speed.
- **RecognitionModeEnum** denotes the recognition modes used during the document or image processing.
- The profiles responsible for the speed results have been updated by the enhanced recognition mode.

- Recognition mode is no longer protected by the license modules, as it was in the previous releases.
- The CLI sample got the new **-rm** key for specifying the recognition mode in the command-line-based applications.

## Engine API

- New **'IsPdfPortfolio'** and **'IsPdfPortfolioFromStream'** methods to find out if an input PDF is a PDF Portfolio.
- **'FREngineDataFolder'** property of the **InitializeEngine** function and method now has new default values. Now, the auxiliary engine data and license data are written into different folders that may be re-defined by a user.
- **'IsPdfWithTextualContent'** and **'IsPdfWithTextualContentFromStream'** methods working with input PDF text layer are now available in Mac version.

## Barcode API

- **'EnableBarcodesCheck'** option of the **BarcodeParams** object specifies whether to use the classifier of the presence of a barcode on an image. Use this property for all barcodes except the postal ones.
- **BarcodeTypeEnum** got the new **BT\_AutodetectWithoutPostal** constant for detecting the type of barcode automatically excluding the postal ones.
- **BarcodeTypeEnum** got the new **BT\_JapanPost** constant which supports the Japanese Post 4-state Customer Code type.

## Document-related API

### IFRDocument:

- **'ConvertFromOldVersion'** - this method loads the document saved only by the previously supported versions of ABBYY FineReader Engine and **FREngineVersionEnum** denotes these previously supported versions.
- **'SourceHasDigitalSignature'** - this property indicates the presence of the digital signature inside one of the documents.
- **'CheckTextLayer'** - this method detects a text on the specified pages or checks the reliability of a text, e.g., the absence of the broken encoding.
- **'AllocatedSize'** - this property returns the size of the memory allocated for the document in bytes.
- **'PDFAttachments'** – this property accesses the collection of the documents extracted from the input PDF files.
- New **'AddImageFileFromAttachments'** method of the **FRDocument** object to open an image file from attachment and add it to a document.

### IFRPage:

- **'SourceHasDigitalSignature'** - this property indicates the presence of the digital signature on at least one of the pages of the source document.
- **'SourceFilePageIndex'** – this property returns the page index in the source document.

### Other changes:

- **PDFAttachment** object was enhanced by the new **'FileFormat'** property to define the image format after its opening in ABBYY FineReader Engine.
- **PDFAttachmentBindingEnum** now has the **PAB\_Portfolio** constant for PDF Portfolios.
- **FontEmbeddingModeEnum** got the new **FEM\_EmbedFullWhenNeeded** and **FEM\_EmbedSubsetWhenNeeded** constants to choose whether the whole font will be embedded into the output PDF or only the subset of a font.
- **FNF\_PDF** constant of **FontNamesFilterEnum** is now available in Mac version.

## Image-related API

- **ImageFileFormatEnum** got the new constants to define the image format when using the **'FileFormat'** of the **PDFAttachments** object: **IFF\_Bmp**, **IFF\_Dcx**, **IFF\_DjVu**, **IFF\_Gif**, **IFF\_Jpeg**, **IFF\_Jpeg2k**, **IFF\_Pcx**, **IFF\_Png**, **IFF\_Tiff**.
- **'IsInMemory'** property of the **ImageDocument** object is now available in Mac version. This property specifies if the image document is stored in memory only or it is also represented as a folder on disk.



## Text-related API

IParagraph object:

- **'UserBookmark'** - this property provides access to a user bookmark by its index in the internal collection of the paragraph's bookmarks.
- **'UserBookmarkCount'** – this property returns the number of user bookmarks in a paragraph.

Other changes:

- Default value of the **'Spacing'** property of the **CharParams** and **FontStyle** objects is now 0.

## BCR synthesis

- **BusinessCardSynthesisParams** – this interface is used for fine-tuning the business card synthesis. This release, it allows you to specify the languages that will be used for processing the recognized text. It is also used as the **'BusinessCardSynthesisParams'** property in the **SynthesisParamsForPage** object.
- **'SynthesizeBusinessCardEx'** – this method tries to detect the business card fields in the recognition area using the specified synthesis parameters.

## Parameter Objects

- **TextLayerInjectionParams** object has got the new **'AllowChangePDFAView'** property for changing the appearance of the output PDF file when facing problems during its processing. It is FALSE by default, so, whenever the document is invalid, an error will be returned, and the document processing will be canceled. If it is set to TRUE, a warning about changing the appearance of the output will be displayed.
- **SourceContentReuseModeEnum** got the new **CRM\_ContentAndPictures** constant which is suitable for PDF files and Office documents with mixed content (text and pictures).
- **'EnableExhaustiveAnalysisMode'** option of the **PageAnalysisParams** object has been marked deprecated and scheduled for deletion in future versions.
- New versions of ALTO (4.0; 4.1; 4.2) are now supported (see **AltoExportParams** and **AltoVersionEnum**).
- **'DetectSpacing'** property of the **FontFormattingDetectionParams** is now FALSE by default.
- New methods for attaching the user-defined pictures to the output PDF:
  - **PDFPicture** interface representing a picture itself and the methods to manipulate the picture position.
  - **PDFPictures** interface representing a collection of the **PDFPicture** interfaces.
  - New **'PDFPictures'** property of the **PDFExportParams** object to access the **PDFPictures** interface.

## License-related API

- **'VolumeRefreshingDate'** method of the **ILicense** object obtains the renewal date of a license with limitations on the number of the pages processed.
- **VRP\_AbsoluteMonth** and **VRP\_AbsoluteYear** constants of **VolumeRefreshingPeriodEnum** support the refreshment of the counter on the specific date.

## Deprecated and deleted functionality

- **'Top', 'Bottom', 'Left', 'Right'** properties of the **BarcodeSymbol** object
- **'EnableAdvancedExtractionMode'** and **'MinRatioToTextHeight'** properties of the **BarcodeParams** object
- **'FastMode'** and **'BalancedMode'** properties of the **RecognizerParams** object
- **Fast Mode** and **Balanced Mode** license modules
- **'SourceHasTextualContent'** property of the **FRDocument** object
- **-fm** and **-rbm** keys for Fast and Balanced modes in the CLI sample
- **'GetEngineObject'** and **'GetEngineObjectEx'** functions for loading the Engine
- **'KeepOriginalCoordinatesInfo'** property of the **PrepareImageMode** object
- **BLT\_Unknown** constant of the **BlockLayerTypeEnum**

- ‘**IsCode39WithoutAsterisk**’ property of the **BarcodeParams** object
- **DIBHello** code sample
- Methods working with Raw bitmap, HBITMAP, DIB (‘**OpenBitmap**’, ‘**OpenBitmapBits**’, ‘**PrepareBitmap**’, ‘**PrepareBitmapBits**’, ‘**OpenDib**’, ‘**PrepareDib**’ methods of the **Engine** object, ‘**EstimateBitmapSize**’ and ‘**GetBitmap**’ methods of the **Image** object, **ITrainingImage::SetBitmapBits**).

## Known issues

This release, the Mac version does not support:

- Office Converters (**OfficeConverters** and **OfficeConversionSettings** objects, **Office Formats Opening** module)
  - Java wrapper
  - Changes in licensing:
    - Activated licenses of Release 4 are irrelevant in Release 5 Update 1.
    - **Licenses** folder is not created after installing Release 5 Update 1.
- Repair method:
- Copy your **Licenses** folder from /Library/Application Support/ABBYY/SDK/12/Licenses to /Users/<user>/Library/Application Support/ABBYY/SDK/12/Licenses.

## Documentation

### Online help

- In addition to the built-in documentation, you may use the latest versions of Developer’s Help and Administrator’s Guide that are publicly available on <https://help.abbyy.com/>. Note that online documentation is protected from indexing.

### Corrected e-mail for activating the license

- Now, to activate your license by e-mail, send the e-mail with the information required for activation to [product-activation-robot@abbyy.com](mailto:product-activation-robot@abbyy.com).

### Better description of Japanese Modern in Recognizing CJK article

- Japanese Modern is a compound language including Japanese and English languages and four letters of the Greek alphabet. Therefore, to obtain the best recognition results for the texts in Japanese, use the Japanese (Modern) language independently and not its combination with the English language.


## Glossary

- The **Introduction** section now includes a new article with the list of terms and abbreviations commonly used all over the text.

## Chinese documentation

- ABBYY FineReader Engine online helps (help.abbyy.com and help.abbyy.cn) include a new guide with GuidedTour section translated to Chinese Simplified. On help.abbyy.com, switch to the articles in Chinese using the language tab.

[Home page](#) → [FineReader Engine 12 for Mac Developer's Guide](#) → [Introduction](#)

English (English)	⬆
Chinese Simplified (简体中文)	⬆

## Unsupported PDF formats

- The **Supported Image Formats** article clarifies what PDF formats are not supported.