ABBYY FlexiCapture 12 Release 3
Release Notes

Table of Contents

Introduction

About this document

This document describes the features that have been implemented in ABBYY FlexiCapture 12 Release 3.

About the product

ABBYY FlexiCapture 12 is the latest version of ABBYY FlexiCapture, a highly scalable and flexible data capture platform for creating region-specific and vertical data capture solutions.

Installing ABBYY FlexiCapture 12

ABBYY FlexiCapture 12 may be installed on the same computer where ABBYY FlexiCapture 10 or 11 is already installed. Please refer to the ABBYY FlexiCapture 12 System Administrator's Guide for installation instructions.

Upgrading from earlier versions

ABBYY FlexiCapture 12 may be installed on the same computer where ABBYY FlexiCapture 10 or 11 is already installed, but you cannot use the setup wizard to upgrade your copy of ABBYY FlexiCapture 10 or 11 to ABBYY FlexiCapture 12.

ABBYY FlexiCapture 12 Release 3 may be used as an upgrade if you have ABBYY FlexiCapture 12 Release 1 and 2 (with any updates) installed on your computer. The earlier release will be removed and Release 3 will be installed instead.

You can use ABBYY FlexiCapture and ABBYY FlexiLayout Studio projects as well as FlexiLayouts created in earlier versions of the program. If you have documents already loaded into the system, we recommend that you first complete the processing of these documents and only then migrate to ABBYY FlexiCapture 12. When you open an ABBYY FlexiCapture or ABBYY FlexiLayout Studio project created in an earlier version of the program, it will be converted to the ABBYY FlexiCapture 12 format. Once a project is converted to the new format, it can no longer be opened in an earlier version.

A detailed description of the upgrade procedure can be found in the System Administrator's Guide.

OCR technologies

ABBYY FlexiCapture 12 Release 3 uses a new version of OCR Technologies (v. 16), delivering a better overall quality of OCR compared to FlexiCapture 12 Release 2 and earlier versions.

See also: *Improved Arabic OCR*, *Improved Japanese OCR*.

Important! Please note that due to changes in the recognition technology, the FlexiLayout matching process in the new version can be slightly different. If a project that you created in an earlier version was designed to accommodate for certain OCR defects, you may now need to change its logic. If you upgrade a project created in an earlier version, please try matching your FlexiLayout on sample documents first.

New hardware requirements for ABBYY FlexiCapture for Invoices

ABBYY FlexiCapture for Invoices now offers advanced field extraction technology powered by deep learning. You will need at least 2.5 GB of RAM per CPU core if you choose to use the thorough field extraction in ABBYY FlexiCapture for Invoices. Please refer to the System Administrator's Guide for more information.

The required amount of RAM will be reduced in future updates to Release 3.

Licensing

For ABBYY FlexiCapture 12 Release 3, you need a serial number generated specifically for ABBYY FlexiCapture 12. Serial numbers generated for ABBYY FlexiCapture 12 Release 1 and 2 (with any update) can also be used for Release 3.

Serial numbers for previous versions of ABBYY FlexiCapture cannot be used for ABBYY FlexiCapture 12.

UI languages

ABBYY FlexiCapture 12 interface is available in the following languages: English, Russian, German, French, Spanish, Korean, Japanese, Czech, Serbian (Latin), Portuguese (Brazil), Polish, Chinese Simplified, and Hungarian.

Notes:

- The web stations are available in all of the above listed languages except Serbian.
- FlexiLayout Studio is only available in English, German, and Russian.
- FormDesigner is only available in English and Russian.
- FCAdminTools is only available in English.

Technical information

Release	Part #	Build #	Installer Build #	OCRT build #	Release date
Release 3	**1299/35**	**12.0.3.2525**	**70734**	**16.1.681.26**	**2019.10.04**

New features in brief

- Invoice processing improvements powered by deep learning
- Extraction of data from unstructured documents using Natural Language Processing (NLP)
- Enterprise readiness enabled by:
    - Performance optimization for small batches
    - Improvements to the backup and restore procedure
    - Command-line interface (CLI) for administering distributed environments
    - Support for Single Sign-On technology

NEW FEATURES

1. ABBYY FlexiCapture for Enterprises

1.1. Optimized performance for small-sized batches: transaction-oriented processing

ABBYY FlexiCapture is optimized to work with batches from ten to several hundred pages in size and can process millions of pages per day (so-called *backend processing*). However, today ABBYY FlexiCapture can also be used as a data extraction service for BPM, RPA, and other systems where ABBYY FlexiCapture receives a flow of small batches that contain only one document with just a few pages each. The end user of the external system will still expect to have his document processed within reasonable time, so the time of processing one document matters and shouldn't increase when the system is scaled up.

In ABBYY FlexiCapture 12 Release 3, the sequencing of requests from the Processing Station has been optimized, resulting in improvements of 50% for small batches (i.e. 3-5 pages per batch) and 10% for typical batches (i.e. 30-300 pages per batch). This means that the system can process millions of pages per day regardless batch size. However, we still do not recommend processing batches of more than 1000 pages.

A detailed description of the configuration used in our benchmark tests is provided in the Performance Guide. The table below is a brief summary of the results.

Configuration	Release 2 Update 8		Release 3	
	B&W pages per 24 hours	Average batch processing time, sec	B&W pages per 24 hours	Average batch processing time, sec
MS SQL, 72 cores, 69 pages per batch	2.0 M	363	2.0 M	368
MS SQL, 94 cores, 3 pages per batch	1.1 M	20	1.7 M	6

1.2. Improvements to the backup and restore procedure

1.2.1. Backup mode (enabled from CMD) to keep data in sync

To ensure uninterrupted operation of the system, the system administrator should schedule regular backups of the data stored in the database and in the file store. For a successful backup, the data stored in the database and in the file store should always be

in sync. At the same time, the system administrator should be able to create backups without stopping the operation of the entire system.

To achieve all of the above, a special mode has been introduced that syncs the data stored in the database and in the file store.

Please refer to the System Administrator's Guide for more details.

1.3. Multitenancy improvements

1.3.1. All Windows stations now fully support dedicated tenant licenses

When several different departments use ABBYY FlexiCapture, dedicated licenses can be issued to track their usage individually.

The Windows Setup, Verification & Remote Verification, Data Verification, and Scanning Stations now fully support dedicated tenant licenses. When a station opens a project on the Application Server, it will know the tenant to which this project belongs and so will use the license corresponding to that tenant.

Working with licenses shared among tenants has not changed.

1.4. CLI for administering distributed environments

1.4.1. Administrators can now use command-line commands to synchronize projects, master data, and training results across multiple ABBYY FlexiCapture installations

Enterprise-level customers typically have several ABBYY FlexiCapture installations: one for development, another for staging, and maybe several more for production in different regions or as a failover backup.

ABBYY FlexiCapture 12 Release 3 offers a command-line interface (CLI) to automate the administration of tasks in distributed environments. The administrator can perform the following using CLI commands:

- Upgrade projects on the Application Server when rolling-out a new version of ABBYY FlexiCapture
- Keep training results in sync across multiple ABBYY FlexiCapture installations
- Transfer project settings from a staging environment to a production environment
- Set up environment variables
- Reuse or upgrade project components, such as Document Definitions and batch types

Please refer to the System Administrator's Guide for more details.

1.5. Platform

1.5.1. Windows Server 2019 support

ABBYY FlexiCapture 12 Release 3 is compatible with Microsoft Windows Server 2019.

1.5.2. PostgreSQL support (beta)

ABBYY FlexiCapture 12 Release 3 can use PostgreSQL as a database to store processing data (instead of Microsoft SQL or Oracle).

We tested the most popular edition, Protgres Pro. Other editions may be tested upon request.

1.5.3. Apache Open Office has been updated to v. 4.1.2

ABBYY FlexiCapture now uses Apache Open Office 4.1.2+ to open office documents. In this version, the known vulnerabilities have been fixed.

1.6. Single Sign-On

1.6.1. Out-of-the-box support of SSO with most popular identity providers and SSO via SAML2.0 and JSON Web Ticket (JWT)

Single sign-on (SSO) enables users to securely authenticate with multiple applications and websites by logging in only once.

ABBYY FlexiCapture 12 Release 3 can be integrated with any identity provider (IdP) using the SAML2.0 protocol or JSON Web Ticket (JWT).

The most popular identity providers, such as Active Directory, Azure Active Directory, OKTA, and OneLogin, are supported out-of-the-box.

Additionally, ABBYY FlexiCapture 12 Release 3 supports SSO with several IdPs simultaneously, and different tenants may have their own unique lists of IdPs (this is necessary when an ABBYY FlexiCapture-based cloud services provided to different customers).

Please refer to the System Administrator's Guide for more details.

1.7. Security

1.7.1. Mutual SSL

ABBYY FlexiCapture fully supports secure communications via TLS1.2 out-of-the-box.

For even greater security of communications, mutual or two-way SSL can be enabled. In this case both the client and the server side will check each other's identity via certification authorities, preventing unverified clients from connecting to the Application Server.

Please refer to the System Administrator's Guide for more details.

1.8. Windows Stations improved

1.8.1. The senior verifier can monitor all tasks at the verification stage, including active, personal and postponed tasks, and take action to comply with the SLA

Some business processes have a strict SLA that defines the time allotted to the processing of documents. Processing steps that require human involvement typically take the most time and may cause delay. Therefore, it is crucial to monitor tasks to make sure that none of them gets stuck at the verification stage. Sometimes an operator may receive a task and postpone it until tomorrow. Or a task may be assigned to an operator that is not available at the moment. The new functionality allows the senior operator to monitor all tasks at the verification stage, whether they are in a queue, assigned to a particular operator, in processing or postponed, and understand their statuses. The senior operator can reassign any queued or postponed task or process. Tasks that are currently being processed cannot be reassigned, but the senior operator will see the username of their respective operators and can either contact them or close their sessions and then reassign the tasks.

This permission is called **View tasks of other operators** in the UI and by default is granted to users with the role of Senior Verification Operator. It can be switched off in the custom role setting. Custom roles can be edited on the **User roles** tab of **Project Properties** dialog box.

1.8.2. Operators of Remote Verification Stations can rerecognize documents locally

The Remote Verification Station is a Windows application that works with projects hosted on the Application Server without using the local licensing service. For ease of administration, the Remote Verification Station does not require the Licensing Server to be installed locally, and so cannot be used to recognize documents locally. However, verification operators sometimes need to change the Document Definition, which requires the documents to be rerecognized. Now, instead of sending the task to server, the documents can be rerecognized locally. Please note that local rerecognition is only available for already recognized pages, as the page counter can only be decreased on the server. As in the earlier versions, pages that have not yet been recognized can only be recognized on the server.

Operators can now also use the **Continue Line Items** command on the Remote Verification Station to extend tables and repeating groups.

1.8.3. For more efficient verification, suggestions with field values garnered through full-text recognition can be switched on or off for each field individually

When a field is not found in the document, the operator needs to enter its value manually. There are several ways to fill in a value quickly. For example, the operator may click on a word on the image or start typing the value in the field. After a document has been recognized, a text layer becomes available at the verification stage. When the verification operator clicks a word on the image, the program will immediately enter this word as the value of the respective field. When the verification operator starts typing a value, a drop-down list appears displaying words from the document's text layer that match the characters typed by the user. This significantly speeds up manual entry of field values.

However, if a field value is hand-printed or hand-written, or is printed with a very rare font, its OCR result may contain too many errors, rendering full-text suggestions useless. In this case full-text suggestions can be switched off for individual fields by opening the properties of the field and disabling the **Use full-text recognition for quick fill-out** option.

With this option switched off, when the verification operator clicks on a word on the image, the respective field will be rerecognized with special field settings and no full-text suggestions will be displayed when the operator starts typing the value manually.

1.8.4. Manual assembly of document sets has been improved based on user feedback

A document set is a way of grouping together documents that belonging to the same case, such as a loan application or an insurance claim. Sometimes the sequence of documents in a document set may be disrupted, or documents from one document set may end up in other documents sets of the same batch. In this case the program will be not able to assemble document sets automatically. To speed up manual document set assembly, the user can create a document set from documents that are out of sequence. The user can select any documents in the batch and apply the **Create Document** command to the selection. The selected documents will then be grouped together into a new document set.

When, for some reason, several documents are merged together, the easiest way to separate them is just to click on a page and make it the first page of new document. In Release 3, this can be easily done by using the **Split Document at Current Page** command.

1.8.5. Duplicating documents to fit complex processing schemes

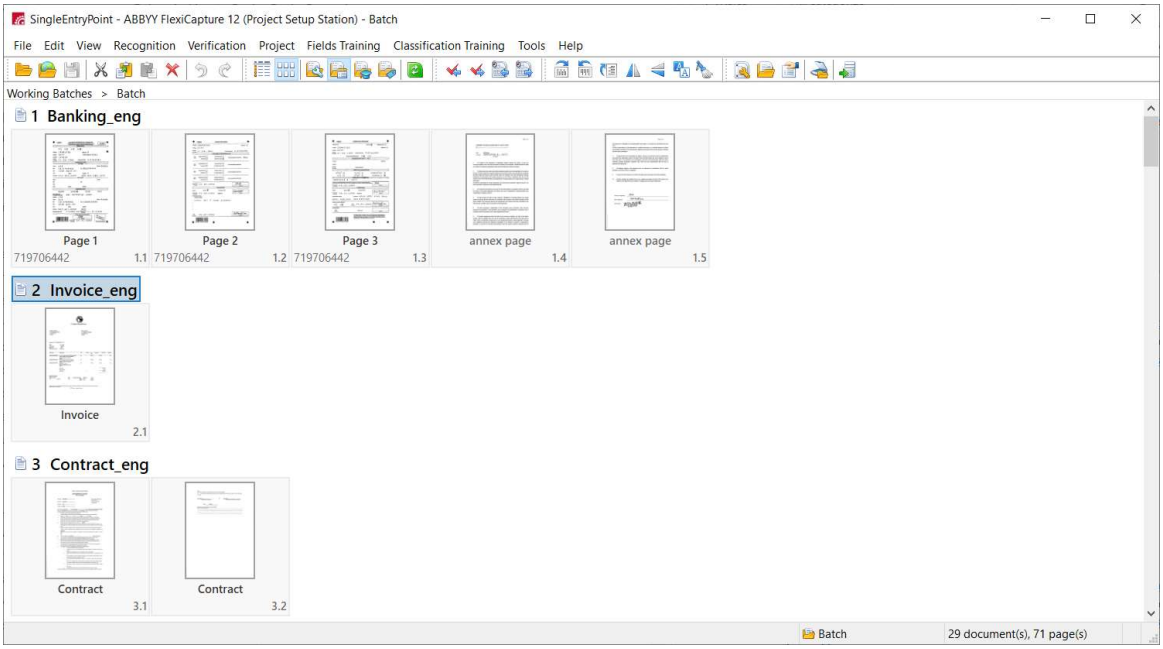
When one physical set of pages represents several documents from a business point of view, it may be necessary to process such documents using different validation rules or send them to different approval workflows. In this case an operator can create a duplicate of an existing document and manually assign another Document Definition to it. A new command, **Tools > Duplicate Documents**, has been added which allows operators create copies of documents within the same batch. This command is available only to operators who have permissions to modify batch content (this can be set up in the custom role settings).

1.8.6. Thumbnail view can now show documents in one or multiple rows to use screen space more efficiently

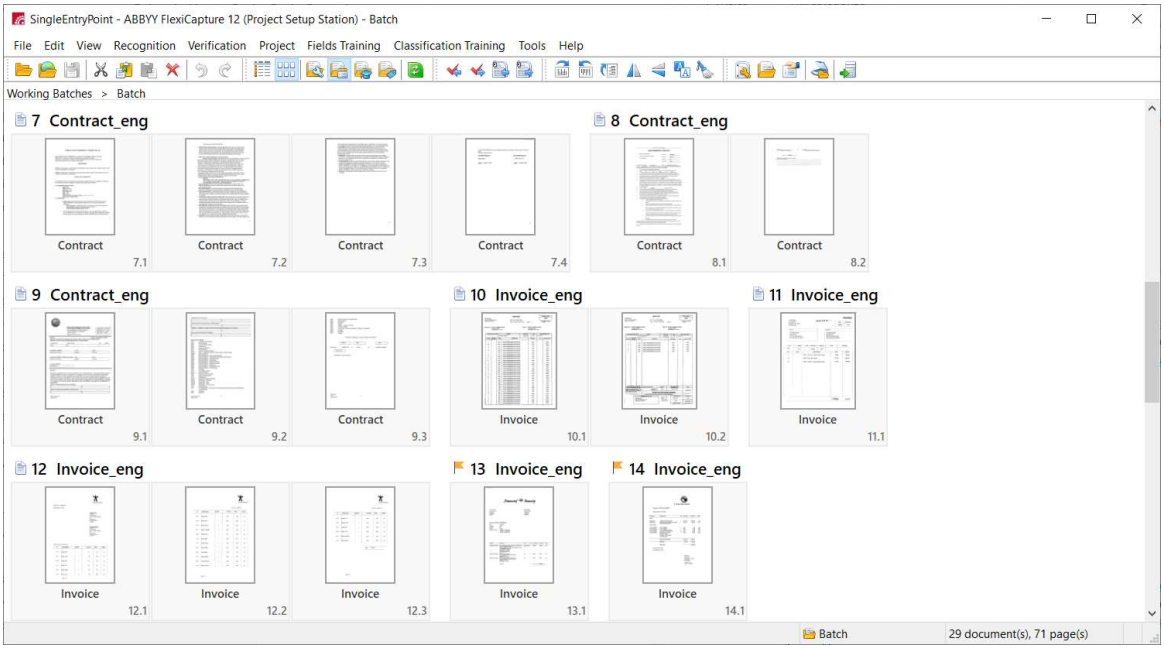
If you work with batches consisting of multiple one-page documents, you may want these documents to take up all the screen space, so that you can see as many documents as possible at once. On the other hand, if you work with multi-page documents, to you may want to start each document on a new line to separate them visually.

To customize thumbnail view, you can now use the **View > Thumbnail options > New Document Starts New Line** command.

Each new document starts a new line:

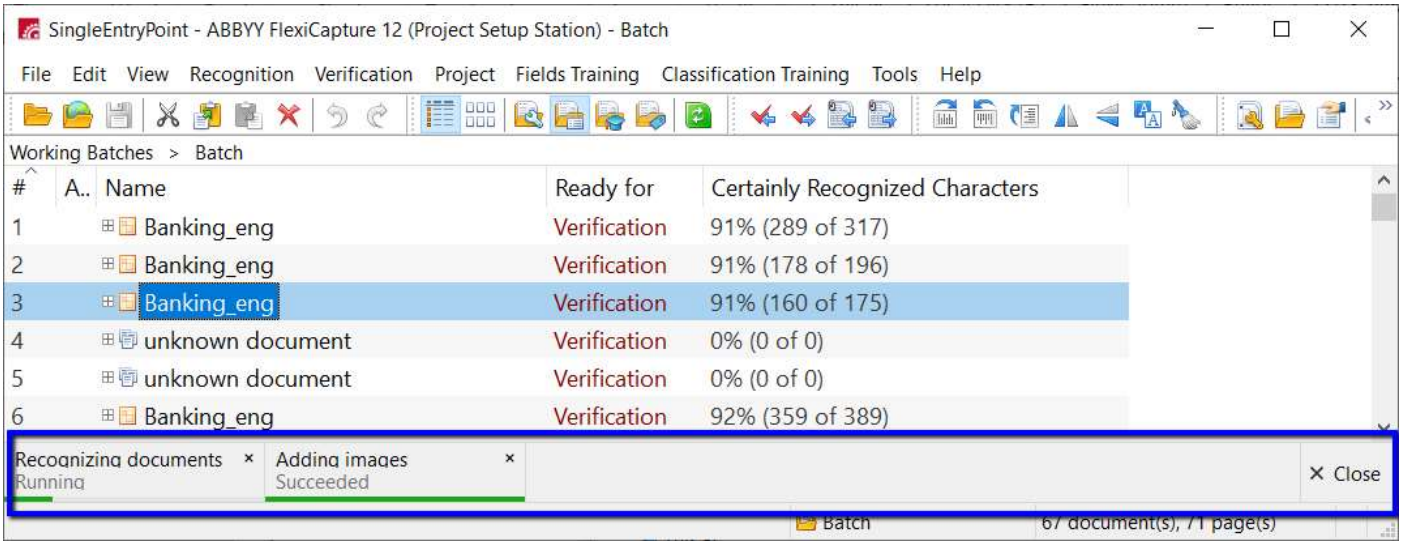


Documents use up all screen space:



1.8.7. A new kind of progress bar has been introduced to speed up operators' work

Instead of showing the progress bar as a separate modal window that prevents the operator from doing other useful work until the current operation completes, a new type of progress bar has been implemented. A new panel at the bottom of the station window will show all the operations running and pending, and the operator will be able to proceed with his work without waiting for the current operations to be finished. For example, changing a Document Definition might involve rerecognition of the documents, which requires some time. Now the user will not need to wait until the rerecognition completes, because the progress bars for all the running tasks will be displayed in a non-modal panel at the bottom of the screen.



To switch the progress panel on or off, use the **Tools > Options > Show Background Task Panel** command. If this panel is switched off, the program will behave as in earlier releases.

2. ABBYY FlexiCapture web applications

Please note that Web Stations based on Silverlight technology are no longer supported and have been removed from FlexiCapture 12 Release 3.

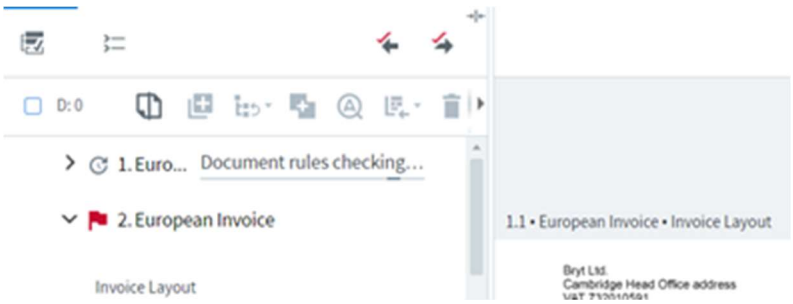
2.1. Web verification improvements

In ABBYY FlexiCapture 12 Release 3, the manual verification process on has been improved.

2.1.1. Rerecognize documents while working with another document in the task

Verification operators sometimes need to make changes to a Document Definition or move pages from one document to another. All these operations will require some of the documents to be rerecognized. Previously, if any documents in a task required rerecognition, the operator had to send the entire task to the server and in the meantime work on another task instead.

Release 3 allows operators to perform server-based operations like recognition or checking rules on one or several documents in a task and continue working with other documents in the same task. Once the server-based operations are completed, the operator will be able to resume his work on the remaining documents. The operator will see a progress panel at the bottom of the screen for documents that are being processed on the server and will be notified when the server-based operations are completed.



2.1.2. Check rules locally

ABBYY FlexiCapture contains a lot of business rules that help validate documents automatically and decide if they are ready to be exported to the backend system. While there are many standard business rules, a typical real-life project will contain a lot of custom rules (e.g. rules implemented using .Net).

Earlier releases of ABBYY FlexiCapture 12 required the majority of the rules to be checked on the server side whenever the operator changed a document.

Starting with ABBYY FlexiCapture 12 Release 3, almost any rule can be checked locally, without contacting the Application Server. This significantly speeds up the verification process. Note that both standard and custom rules created using scripts can be checked locally. While there are some obvious limitations (e.g. you cannot access the file system of a Processing Station from a script rule), generally, you need to create your script rules only once. Your rules will be suitable for checking both on the Processing Station and on the Verification Station.

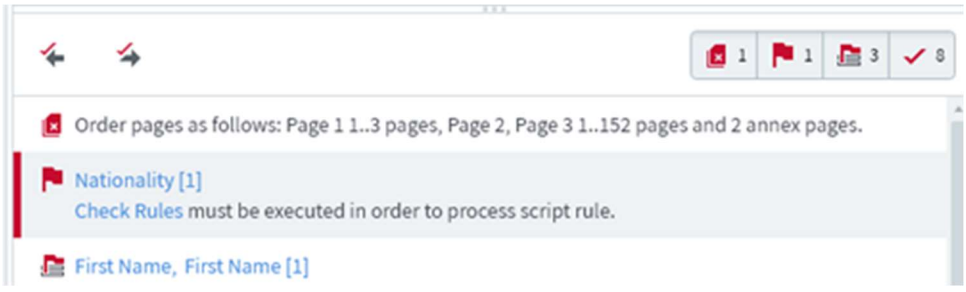
For your script to be executed without calling rules, the objects used in the script should not require rules to be checked on Web Stations. If your script contains at least one object that requires rules to be checked on a Web Station, the “Check Rules” button will be used for checking rules. Detailed information can be found in the description of each object available here: http://help.abbyy.com/en-us/flexicapture/12/distributed_administrator/objects.

Rules are checked on machines where the Verification Station web application is running. If you have a large number of rules, be sure to monitor the workload of these servers. You may need to add more servers to the NLB cluster containing the Verification Station web application if you have hundreds of verifiers working simultaneously.

Note that the following are still not supported by the Verification Station:

- normalization of field values
- “sum in figures - sum in words” rule

For this reason, the “Check Rules” button is still available in the GUI. Clicking this button will cause rules to be checked on the Application Server, in the same manner as rules are checked on the Processing Station. The screenshot below shows an example of an error message generated by a script rule that uses an unsupported interface that is currently not available on the Verification Station for local checks).



This limitation will be removed in a future update to Release 3.

The checking of script rules on the Verification Station is enabled by default. In the event of any issues, you can always revert to the behavior of the earlier releases, where all rules are checked only on the Application Server. To disable the checking of rules on the Verification Station, open the *web.config* file and set the *DisableScriptRules* key to "true."

2.1.3. Faster queues view

The queues view shows the number of tasks in different queues. Previously, it took quite some time to collect all the data about all of the queues when there was a large number of tasks in the system. In Release 3, the display of queues has been optimized and now 10+ queues with several thousand tasks take less than 1 second to load.

2.1.4. Faster task view

The Web Verification Station has been reengineered using the modern React/Redux stack technology. This has allowed us to optimize the opening of tasks. A new task should now open within 3-5 seconds regardless of how many documents it contains.

The list of page thumbnails also scrolls and loads faster, regardless of the number of pages in the document.

The data form has also been optimized to work faster with documents that have a lot of fields (e.g. thousands of line items within one multipage invoice).

Note that the responsiveness of the task view on the Web Verification Station now depends only on the client hardware.

The recommended system requirements are:

- Intel® Core™ i3 processor or equivalent
- 4 GB RAM

2.2. Invoice processing improvements

2.2.1. Vendor and BU lookup works both when the vendor depends on the BU and when there is no such dependence

In the case of invoice processing, master data may contain either independent lists of vendors and business units (i.e. any vendor can send an invoice to any business unit), or vendors may depend on business units (this typically happens when each business unit has his own master data). If vendors depend on business units, the **BUId** column should be filled in the "Vendors" data set.

Vendor and BU lookup on the Web Verification Station will now work properly in either case, showing only the vendors that are allowed for the current business unit, depending on how the master data are configured.

2.2.2. Purchase order matching implemented

Purchase order matching allows matching invoice line items to their corresponding order items in the data set. Now this functionality is fully supported by the Web Verification Station.

The operator can see a list of purchase order items by clicking the **Details** button next to the **Purchase Order Number** field on the invoice data form. This list provides some advanced capabilities to simplify matching order and invoice items, such as search and the ability to rearrange columns.

2.2.3. UX improvements for line Item verification

Based on feedback from users, a number of UX improvements have been made to speed up verification of line items.

- Users can now navigate inside line items using the "Up," "Down," "Left," and "Right" arrow keys on the keyboard. To start navigating a line item, shift the focus to the line item field. If there are two or more repeating groups inside the line item field, each group will be treated as a separate unit.
- The **Continue Line Items** and **Reanalyze** commands are now additionally available to operators in the following cases:
    - entire row is in focus
    - focus is on a cell and at least one cell is filled in the row containing the cell
- If an operator deletes all rows on the data form, all the regions on all the pages will be automatically deleted for the repeating group in the current line item.

2.3. Other improvements

2.3.1. Page numbering in tasks on Web Verification Stations

For easier navigation between documents and pages, pages are now numbered in the batch editing and document editing portions of the screen.

2.3.2. Faster and more stable resizing of window panes

The mechanism that redraws pane separators has been improved for faster and more stable resizing in all supported browsers.

2.3.3. Senior Verification Operator is automatically returned to the original queue after (s)he finishes working with a task

When the Senior Verification Operator finishes working with a task from the task list, whatever the result, the operator will be returned to the same queue if it still contains unverified tasks. Otherwise, the operator will be redirected to the list of all queues.

2.4. Localization of the data form on the Web Verification Station

The localization of the data form depending on the operator's language is now fully supported on the Web Verification Station. The localization of the data form can be configured on the Project Setup Station in the Document Definition editor by clicking **Document Definition > Localization**. Please refer to the User's Guide to learn how you can configure the localization of field captions, rule messages and, other document elements.

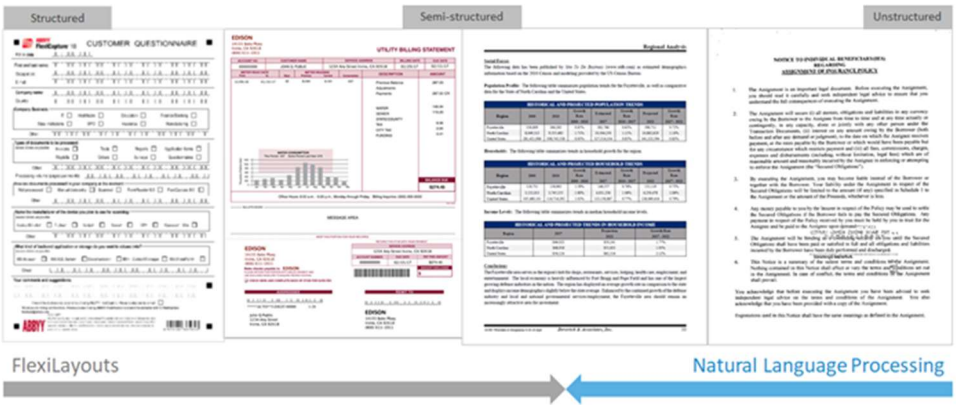
2.5. Ability to download exported results on the Web Verification Station

In the case of self-service, document processing results can be downloaded onto the Web Verification Station once the document has gone through the export stage. The download feature is only available to the senior verifier and is disabled by default. To enable this feature, open the *web.config* file of the Web Verification Station and set the *CanDownloadBatchResult* key to "true." The date and time of the last download will be stored in a batch registration parameter called "BatchDownloadTime."

3. Data extraction using Natural Language Processing (NLP) technology

3.1. What documents can be processed with NLP?

FlexiCapture traditionally focus on processing structured documents, such as tax forms, and semi-structured documents, such as invoices and bills of lading. **NLP technology, available in Release 3, extends capture capabilities to a new content type—unstructured documents (e.g., contracts, leases, articles, agreements, email).**



Structured documents. In the case of structured documents, certain types of information always appear in the same places on the page. Good examples of structured documents are census forms, questionnaires, and loan applications.

Semi-structured documents. If you examine a few invoices from different vendors, the data will seem to be positioned randomly on the page. On the other hand, if you examine several invoices from the same vendor, you will notice that there is some structure. If you examine a sufficiently large number of invoices, you will have to conclude that while the positioning of the fields differs a lot from invoice to invoice, there are still some typical geometric areas where certain types of information can be found. For example, vendor's name is usually located somewhere at the top of the page, invoice number and date are located somewhere at the beginning of the document. Documents with such geometrical arrangement of data are termed "semi-structured."

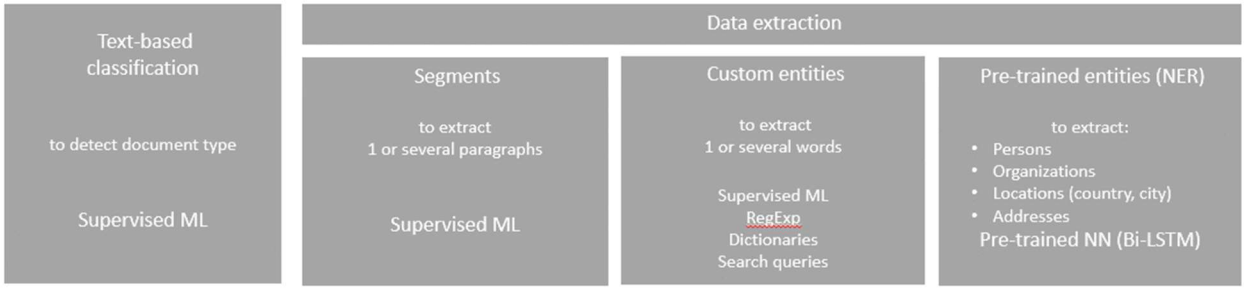
Unstructured documents. Finally, if you examine some contract, you will see that geometry is of no help to locate the parties or the amounts, as the document mainly consists of the text and is completely unstructured geometrically. But lawyers familiar with the rules and principles of contract drafting will always identify the structure hidden in the contract and will easily locate the required information. ABBYY FlexiCapture does the same relying on Natural Language Processing (NLP) technology – it can "read" the text, "understand" its meaning, and extract required information.

Combination of semi-structured and unstructured information. There are also lots of documents, such as appraisal reports, that contain both semi-structured portions like tables and unstructured portions made up of continuous text. In this case, a

combination of different technologies should be applied. Using FlexiLayouts, you can capture a data from tables, and use NLP to extract data from text.

3.2. What data can be extracted with NLP?

ABBYY FlexiCapture allows you to train classification, segmentation, and entity extraction machine learning models to capture data specifically required by your business processes.



Classification

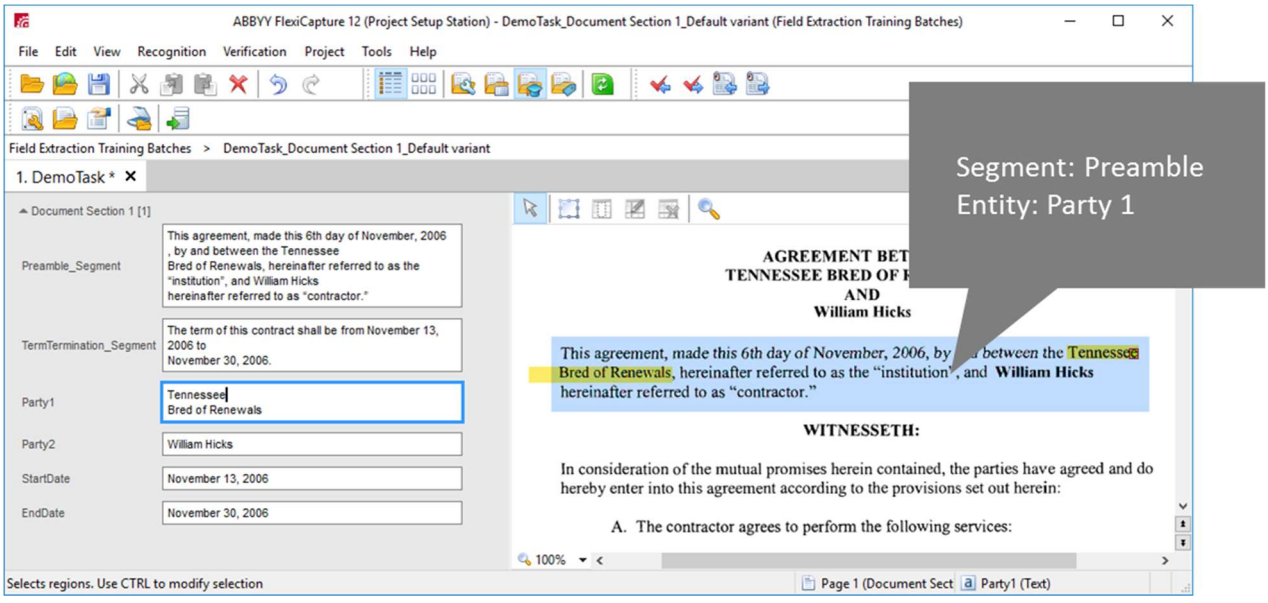
To extract data from a document, you typically first need to identify the type of your document. ABBYY FlexiCapture contains advanced classification technology that uses both image and text based[1] features to identify document class. Document classification enables the program to determine which Document Definition should be applied for further analysis. Classification has been available in ABBYY FlexiCapture 12 starting from Release 1. Please refer the System Administrator's Guide for detailed information about document classification.

Segmentation

Once the program knows the document type, it goes deeper and tries to locate the paragraph that most probably contains the required information – just like a human, who quickly looks through a multipage document trying to find, for example, the Terms of Payment section. This process is called document segmentation. It retrieves a relevant text segment that can be further analyzed more thoroughly. Segmentation can be also used for clause detection in contracts.

Entities extraction

Once the required paragraph is identified, the human will read it closely to "extract" the necessary information. ABBYY FlexiCapture does the same. Using information from NLP parser, ABBYY FlexiCapture extracts the syntactic and semantic structure of the sentence, determines the part of speech that each word belongs to, identifies their morphology, syntactic and semantic roles, and semantic classes. This sophisticated knowledge is used as features in machine learning model that extracts custom entities from the text.

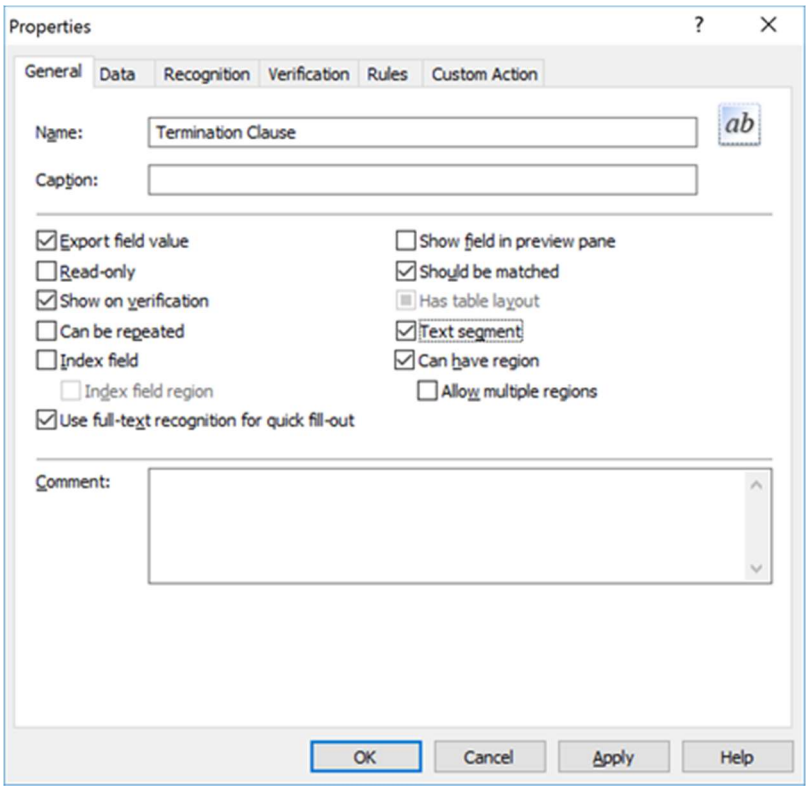


[1] Text-based classification is usually called NLP classification

3.3. How to train NLP Machine Learning model?

To be able to use the segmentation and extraction technologies to process unstructured documents, you need to install the ABBYY FlexiCapture NLP add-on. The NLP add-on is a separate installation package. When installed, you will see an "ABBYY FlexiCapture NLP" item in the Windows Programs and Features window. This component should be installed on all the machines where either training or processing occurs, i.e. on the Processing Stations, Project Setup Stations, and the Windows Verification Stations. Additionally, the NLP option should be enabled in your license (this can be checked in the License Manager).

To train segmentation and extraction, you first need to create fields in your Document Definition for the data you are going to capture. Segments are text paragraphs and may contain multiple fields. Open the **Properties** dialog box for a field and select **Text segment** option on the **General** tab. Now when working in the document editor, you will be able to draw regions for child fields inside the segment's region.



Then you need to declare the NLP models that should be trained. In the **Document Section Properties** dialog box, click **NLP** tab. Here you can either load NLP models that have already been developed by ABBYY engineers or create your own. When creating your own NLP model, you will need to define the following:

- Type. This can be either segmentation (finds a text fragment inside a document) or extraction (finds specific entities inside a text fragment).
- Source. Segmentation always uses the text of the entire section to find the required paragraphs, while extraction models can deal with text from different fields, located either by using segmentation (e.g. a clause of a contract) or by using FlexiLayouts (e.g. an invoice line item).
- The language of the text to be analyzed.
- A set of output fields. These may be a field for the segment or fields for contract parties and amounts.

You can create several NLP models. We recommend creating one segmentation model for each section and then one extraction model for each source field, e.g. for each segment.

Once you have finished creating your models, you need to create a training markup. Proceed to **Field Extraction Training Batches**, create a batch and connect it to the Document Definition Section with your NLP models. On the shortcut menu of the batch, enable the **NLP batch** option (this will tell the program that it should train NLP models and not FlexiLayouts).

Load documents into the training batch, set the status of the documents to "For Training," recognize the documents, and create regions for all the fields that should be extracted by NLP. Make sure your markup is consistent throughout. Observe the following guidelines:

- Markup everything you place into your training batch. If you set the status of a document to “For Training” and this document contains the fields that you are looking for, be sure to mark up the regions of these fields, otherwise the program will think that there are no fields to extract there.
- Indicate each required field in the same place on all documents. If a required field appears in different positions on a document, stick to the position that is common to all documents.

To achieve the best results with NLP, do not try to teach the program the business logic of your documents. NLP is great at understanding sentences, it can differentiate buyers from sellers or start dates from finish dates, it can even determine if a headache is a symptom that requires taking a medication or a side effect after taking a medication. But consider the following simple example. Suppose you need to capture the termination date of a contract. Looking through the documents, you discover that sometimes a specific date is mentioned as the termination date (e.g. “12/31/2020”), while in plenty of other cases you can only find a period of time or duration (e.g. “3 years from the Effective Date.”). To achieve the best extraction quality in such cases, we recommend creating and training separate fields for NLP extraction: “end date” and “duration.” Then you can create a business rule that will calculate the end date by adding a duration to the start date. NLP can learn very well what it should find and extract from a text. But it is by no means an expert and cannot make its own conclusions. For example, it will not know how to deduce the end date from a start date and duration. You will need to create business rules to populate your data model with data based on the information extracted by using NLP.

NLP technology will benefit from large numbers of documents, even though it can start working even if it is given only a few samples. Please note the more documents you have, the longer the training will take. The best approach is to provide NLP with a representative training set. If you are training a lot of fields, the variations of each field should be represented. You can start with a dozen documents, check the extraction quality, and add more samples if you see that some of the fields have not been located properly.

Once you have marked up the documents in the training batch, click **Train** (or press Ctrl+F7) and wait until training is finished. Training times depend on the number of samples, the number of models, and the number of fields.

You can then load new documents into the working batch and see how well segmentation and extraction perform.

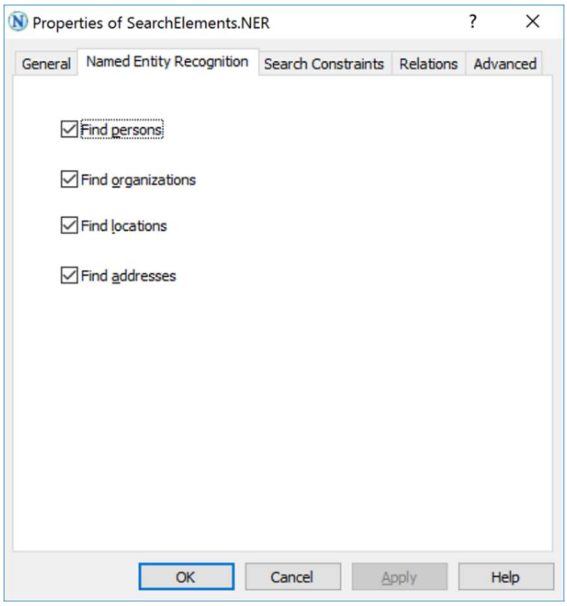
3.4. What if you do not have enough documents for training? NER (Named Entity Recognition), RegExp (regular expressions), dictionaries, search queries.

You may need to correct NLP results when you do not have enough documents to train machine learning model.

Segmentation

You can use a FlexiLayout instead of segmentation. For example, you can set up a FlexiLayout to look for some paragraph by iterating through different paragraphs in the text. You can then decide if a paragraph is the right one by looking at its title, by examining some words or entities inside, or by analyzing the contents of the neighboring paragraphs.

Named Entity Recognition (NER) has been added to ABBYY FlexiLayout Studio in Release 3. A new search element, called “Named Entity,” allows you to locate persons, organizations, locations (e.g. cities or countries) and addresses. Note that this technology is optimized to locate entities inside natural text that is made up of sentences.



Using rules combined with NER, you can now tell the program, for example, that you are looking for paragraphs where at least two persons are mentioned.

Entity extraction

You can use extraction scripts instead of extraction models. Extraction scripts can be created in the Document Definition editor on the Project Setup Station. They work with the raw text of the source field and do not know anything about the positions of objects on the page.

An extraction script first needs to know what kind of entities you are looking for. It can capture entities using NER and it can capture words using regular expressions and a user-defined dictionary. Then you can specify, for example, that you want to populate the field with a person (i.e. a set of words detected as a person by NER) that is closest to the word "buyer" from a user-defined dictionary (typically, the dictionary will contain all the inflected forms).

3.5. Controlled launch process. Proven and new cases

NLP capability will be generally available but in a controlled release. What is the difference between Controlled launch and standard FC deals?

1. During Control Launch Process, all NLP models are trained by ABBYY Professional Service engineers, not by customer.
2. We split all NLP cases into 2 groups: proven cases and new cases.
    a. **For Proven cases** no approval is needed. List of proven cases:
        - data extraction from Commercial Leases
        - data extraction from Loans for Commercial Real Estate

    b. **New NLP cases**. All customer and partner deals will require all sales to go through a review and approval process.
3. **NLP projects requirements.** If you have an NLP case, please provide following info to your ABBYY sales representative:
    a. Use case overview
        i. What is customer company profile?
        ii. Describe in 1-2 sentences what business problem customer would like to solve, and how they are solving it now.
        iii. Timeframe to complete the project
        iv. How many documents (pages) should be processed? Is it a one-time project or ongoing process?
    b. What docs should be processed? What fields should be extracted?
        i. Please provide List of fields that should be extracted in Excel file.
        ii. Marked-up 2-5 samples in PDF: highlight the fields & add field name (as in Excel file) as a comment
        iii. provide 100+ samples (without markup) to train ML model

3.6. Languages

NLP data extraction can be done for English, German, and Russian languages.

3.7. NLP Help

User Guide provides detailed information on how extract data using NLP.

4. ABBYY FlexiCapture for Invoices

4.1. Deep learning for invoice header fields and line items

A new technology based on neural networks is now used to extract header fields (e.g. invoice number, invoice date, and total amount) and line items from invoices. The out-of-the-box quality of field detection has significantly increased. Benchmark tests where Release 3 was compared against Release 2 showed that:

- Detection of **Invoice Number** and **Invoice Date** field has improved by 5-10% and now stands at **93+%**.
- Detection of the **Total** field has improved by 15-25% and has reached **75+%**.
- **75+%** of **line items** are now found (improvement of 8%) and for **55%+ of line items** all key fields are now found (e.g. “unit price,” “quantity,” “total”).

Invoices from the countries listed below have been used to train ABBYY FlexiCapture.

For header fields: Austria, Australia, Belgium, Canada, the Czech Republic, France, Germany, the UK, Greece, Hungary, Ireland, Italy, the Netherlands, New Zealand, Poland, Singapore, Spain, Sweden, Switzerland, and the USA.

For line Items: Austria, Canada, France, Germany, the UK, the Netherlands, Switzerland, New Zealand, and the USA.

New trained fields and supported countries will be added in future updates.

The table below shows the quality of extraction for header fields and line items for the main supported countries.

Country	Invoice Count	Invoice Number		Invoice Date		Total		Line Items			
								Rows		Quantity, Unit Price, Total	
		F	*Delta to R2*	**F**	*Delta to R2*	**F**	*Delta to R2*	**F**	*Delta to R2*	**F**	*Delta to R2*
AU	2726	**95.58%**	*17.82%*	**92.88%**	*4.70%*	**68.44%**	*6.35%*	**81.85%**	*18.11%*	**63.38%**	*18.33%*
CA	638	**91.70%**	*10.13%*	**90.06%**	*7.91%*	**83.76%**	*9.82%*	**80.53%**	*4.49%*	**61.73%**	*9.61%*
DE	491	**94.04%**	*26.99%*	**92.34%**	*23.02%*	**92.86%**	*23.51%*	**62.25%**	*12.50%*	**48.71%**	*10.33%*
FR	568	**89.95%**	*6.13%*	**90.10%**	*3.11%*	**68.29%**	*11.00%*	**67.98%**	*3.42%*	**52.94%**	*2.76%*
GB	646	**95.81%**	*3.42%*	**95.39%**	*7.12%*	**83.82%**	*8.19%*	**80.15%**	*12.55%*	**64.41%**	*11.49%*
SE	628	**93.89%**	*5.05%*	**95.06%**	*5.78%*	**56.67%**	*24.71%*	**68.09%**	*17.55%*	**55.90%**	*15.01%*
US	3265	**88.32%**	*5.69%*	**88.33%**	*2.99%*	**74.97%**	*11.26%*	**77.62%**	*4.64%*	**45.21%**	*5.60%*
Average	8471	**92.76%**	10.75%	**92.02%**	5.38%	**75.54%**	11.89%	**74.07%**	10.47%	**56.04%**	10.45%

Notes:

- As the new technology requires more computing resources, we recommend installing the x64 version of the program and having at least 2.5 GB of RAM for each CPU core (i.e. for each FlexiEx.exe process). Note, however, that using this technology may still increase the processing of each page by about 1 second.
- The new technology is geared specifically toward invoice processing and is not recommended for use on other documents, even if their layout is similar to that of invoices (e.g. purchase orders or waybills). The new technology can

be switched off in **Document Definition Properties** > **Invoice Settings** > **Additional Fields and Features** > **Thorough extraction of invoice header fields** for header fields and in **Document Definition Properties** > **Invoice Settings** > **Additional Fields and Features** > **Thorough extraction of invoice line items** for line items.

4.2. Improved machine learning for line items

Machine learning for line items has been improved both for invoices and for general document processing.

- The data type of data in each column is now taken into account, which improves the quality of column identification and prevents erroneous capture of neighboring columns.
- Detection of line item borders has been improved, preventing erroneous detection of line items in the header and footer of a page. The program learns how text rows containing the line items look and how to differentiate them from the text rows containing the header and the footer. For this reason, it is important to provide a complete markup of all the line items when you train the program. If your document contains a very large number of pages, you can provide just a few pages, and delete the rest. For example, if you have a 100-page invoice, it may well be sufficient to provide the first, the last, and a middle page, as the other 97 pages will most likely look the same as the middle page you have selected. But for all the pages that you provide to the program, the markup should be consistent, otherwise the program will be unable to differentiate text rows containing line items from other rows.
- Hidden page classification has been trained and applied to prevent false detection of tables on pages that actually do not contain tables. For example, in the case of a bank statement, a table can start on the second page. To prevent false table detection on the first page, ABBYY FlexiCapture will automatically learns how to classify pages some of which contain a table.

4.3. Machine learning for processing invoices with different layouts from the same vendor

When one vendor issues invoices with different layouts, training one FlexiLayout for all possible layouts from that vendor will not produce very good results.

ABBYY FlexiCapture 12 Release 3 offers clustering technology to address the issue. When a new document arrives, its vendor is detected first. Based on the vendor, the special training batch is selected that stores the field extraction model for this particular vendor. If the program fails to detect the vendor, the “unknown vendor” training batch is selected. Training is done separately for each training batch, but instead of creating one FlexiLayout for all the sample invoices in a training batch, the program will first cluster the samples (i.e. group all similar-looking documents together) and then train a separate field extraction model for each cluster. This approach yields high-quality results even if a batch contains documents of varying layouts.

The clustering is not directly accessible to operators. Operators will simply have to indicate the locations of the fields during training and the program will cluster similar-looking documents automatically.

You can enable the clustering feature in **Document Definition Properties > Invoice Settings > Additional Fields and Features**.

When upgrading from earlier versions, FlexiLayouts that have been already trained or manually developed for specific vendors can be used without any changes. But when the user starts a new training session, invoices in his training batches will be clustered and a new set of FlexiLayouts will be created for each cluster. Locked and manually created FlexiLayouts will remain unmodified.

If you are not satisfied with your training results for a vendor, you can create a FlexiLayout for this particular vendor manually. In this case, you will still create one FlexiLayout per one training batch and the documents in this training batch will not be clustered. Therefore, you cannot manually create a FlexiLayout for each cluster. Clustering is an integral part of training and is optimized to create clusters with field training in mind. It is always done automatically, as the algorithms involved are too complicated and cannot be easily fine-tuned by the user.

4.4. Seamless upgrades of invoice projects

A new upgrade procedure protects users from losing their custom settings, because no custom rules or fields will be affected by the upgrade. Only new rules and fields can be added during an upgrade, provided they were added to the supplied Document Definition. Any changes to the rules and field properties made by the user will be preserved.

4.5. Tax extraction can now be switched on and off

Now you can switch tax extraction on or off as an invoice feature in **Document Definition Properties > Invoice Settings > Additional Fields and Features**. All the related fields and rules will be switched off or on automatically.

5. General improvements

5.1. Image enhancement

5.1.1. Image enhancement profiles can be applied manually on the Windows Verification Station

Image enhancement profiles can be applied automatically after loading images into a batch, depending on the batch type or image processing settings. These profiles can also be applied by scripts. Different profiles can be applied to images of different types.

If a profile is applied to the wrong image, or if a profile worsens rather than improves the result, the operator can revert to the original image and apply another profile.

To apply a profile manually to already uploaded images, open their shortcut menu and click **Edit Image > Apply Image Enhancement Profile**.

5.1.2. Ability to view the original image during verification and to export it

If an image has been enhanced to improve the quality of OCR, the original image can be stored and accessed if required.

- The original image can be displayed during verification to check if any information has been lost, if autocrop has worked as it should have, or to see the deleted stamps and signatures.
- The original image can be exported for archiving purposes.
- The user can always revert back to the original image using the UI or scripts if he is not satisfied with the quality of the enhanced image or if the profile has been applied to the wrong image by mistake.

Original images will be saved if the **Store original image during processing** option is selected. Storing original images will increase the size of the file store because two images will be stored for each page – the original image and an enhanced image.

To see the original image, open the shortcut menu and click **Edit Image > Show Original Image**. To export the original, select the **Use original image** option when setting up an export profile. To revert to the original image, click **Edit Image > Revert to Original**.

You can use scripts to analyze the suitability of an image for OCR after various image enhancement profiles have been applied. If you want to revert to the original image, use the *IPage.RevertToOriginalImage()* method. Reverting to the original image is only possible if the original image has actually been stored for the given page (as indicated by the property *IPage.HasOriginalImage* property). For more information, please see http://help.abbyy.com.

5.1.3. New image enhancement operations are available in scripts

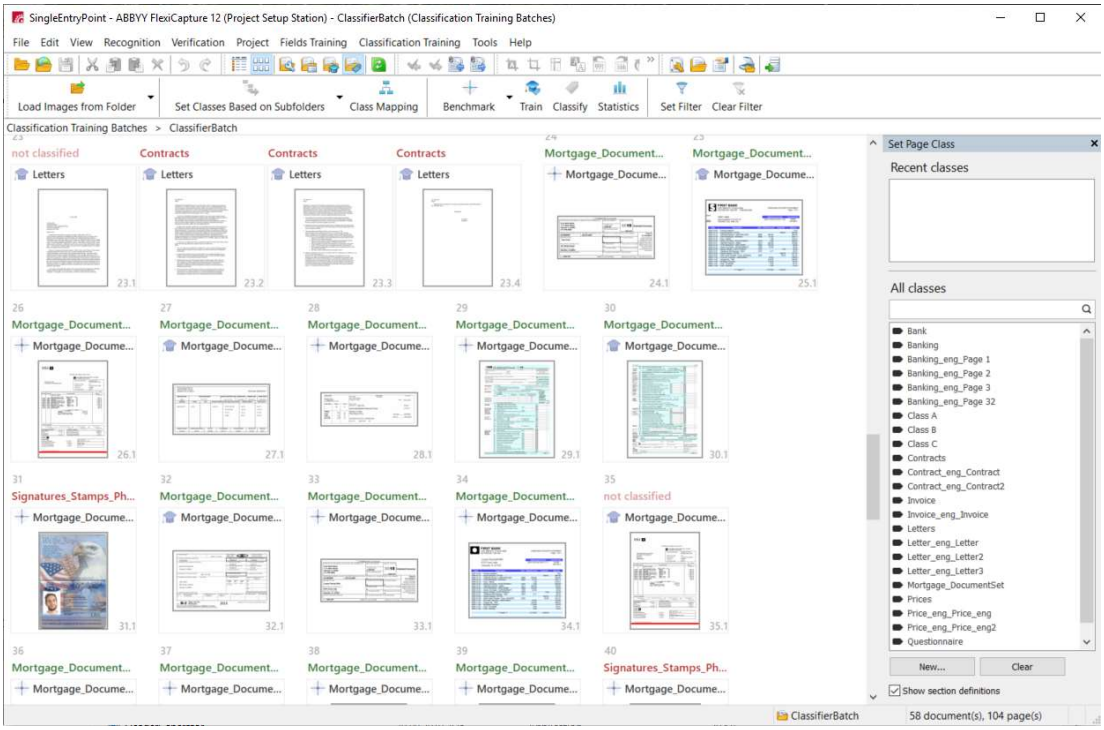
New methods have been added to *IeditablePictureObject,* which allow you to apply any image enhancement profile from within a script.

- **AutoCrop([optional] type: string)**; // Values of type "Photo," "Scan," and "Generic" are allowed. An empty value type is treated as Generic.
- **ApplyMultiscaleLocalContrastFilter()**; // increases contrast
- **RemoveMotionBlur()**; // removes motion blur

For more information, please see http://help.abbyy.com.

5.2. Ability to train classifiers directly in ABBYY FlexiCapture

5.2.1. Overview



ABBYY FlexiCapture includes a powerful classification technology that can identify document types, assemble documents, or detect the topic of an article or the intent of an e-mail message. Classification can be trained and then improved on-the-fly by using feedback from the verification stage. The classification designer offered by the Project Setup Station has been significantly improved in Release 3 to allow advanced users to train and fine-tune classification using samples accumulated during production. All this can be done directly in ABBYY FlexiCapture, without starting ABBYY FlexiLayout Studio.

To setup a classifier, you need to complete the following steps:

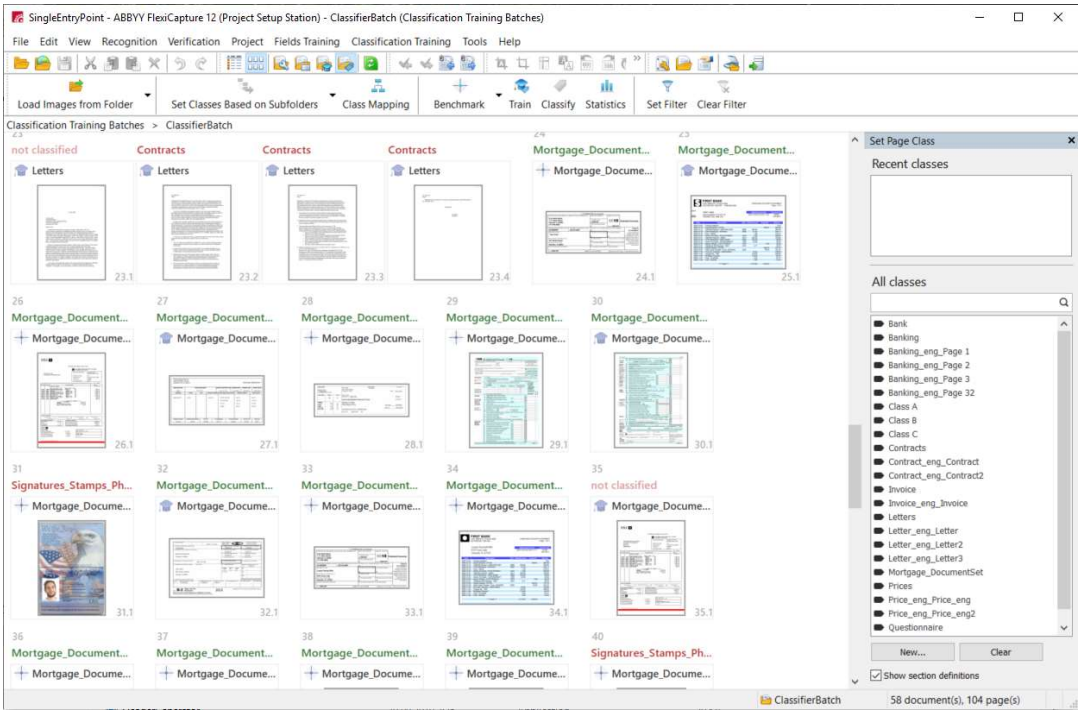
1. Import documents and assign reference classes to them.
2. Train the classifier and analyze any issues.
3. Improve the training set accordingly.
4. Establish correspondences between classes, Document Definitions, section variants, and page roles (i.e. the first page, any page).
5. Publish the classifier for the batch type.

5.2.2. Import a training set and assign classes automatically according to folder names

If your customer has already sorted their documents into classes and put them into their appropriate folders, ABBYY FlexiCapture can import documents form these folders and assign classes based on their names. Use the **Load Images from Folders** command to import all images at once, and then use the **Set Reference Class by Source Folder** command to assign a reference class based on the name of its folder. It is always a good idea to double-check that customer has sorted their documents properly.

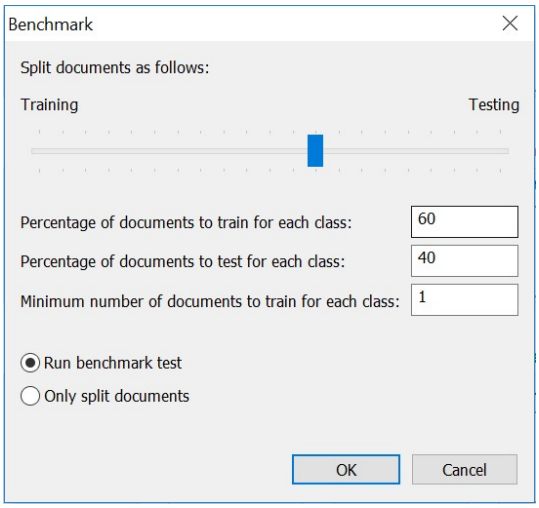
5.2.3. Thumbnail view optimized for handling classification batches; a special panel added for setting page classes quickly

The thumbnail view for classification packages has been optimized to make setting up and improving the classifier more convenient. In thumbnail view, each page contains information about its reference class, result class, and document status (i.e. “for training,” “for testing,” or “unused”). For easier class assignment, a **Set Page Class** panel has been added, where all the used classes and Document Definition sections are displayed. In this panel, you can search for classes or Document Definition sections and create or delete classes.



5.2.4. Automatically split a batch into training and test sets and run a benchmark test

In order to evaluate the quality of your classifier, you need to divide your set of documents into two parts — one will be used for training and the other for testing. The program will learn how to classify your documents using the training set, and the quality of classification will be tested on the test set. The benchmark feature allows you to split your batch of documents automatically – all you need to do is specify the ratio of training documents to test documents.



5.2.5. Tools for analyzing classification statistics: confusion matrix and confusing classes

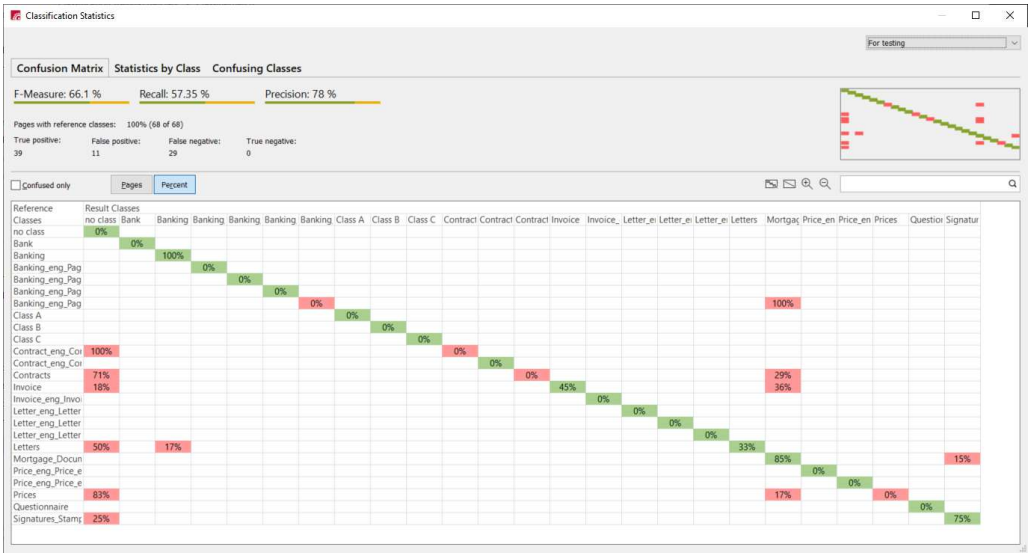
When the benchmark test is finished, statistics are shown for the test set. In addition to the main quality indicators, such as F-measure, Recall and Precision, other statistics are presented on the following three tabs:

Confusion Matrix. This visual tool helps you to figure out which classes are sometimes confused by the classifier. The rows of the table represent the reference classes and the columns represent the result classes. The diagonally running green cells contain the correctly classified samples, and everything outside are classification errors. Clicking on any cell will display the documents where the classification error occurred. Use the **Show Similar** tool to investigate the training set and understand which sample documents may have caused this particular classification error.

Confusing Classes. This tab contains a list of classes that have been mixed up by the classifier. This is the same data as in the confusion matrix, but it is represented in list form. This representation allows you to sort out issues when you have a large number of classes, focusing only on those that will contribute the most to the overall quality of classification.

Statistics by Class. This tab contains statistics for each class and helps you identify classes with the lowest quality. You will probably have to improve the training set for these classes.

You can also review statistics for the training set and for all the images in the classifier batch. Use the drop-down list in the top right corner of the **Classification Statistics** window to switch between the different types of statistical data. Analyzing classification errors in the training set will help you find incorrect markup.



5.2.6. If a classification error occurs, you can search for similar pages in the training set to identify markup issues

ABBYY FlexiCapture technology is based on state-of-the-art CNN and RNN and can produce excellent results even in very complicated cases. In the majority of cases, the quality of classification can be improved by fixing markup errors in the training set.

Even if you mark up your documents yourself with a lot of concentration, after marking up several hundred documents you will most likely start making errors. Typically, a markup created by one person contains about 3% errors. This may be good to know, but finding these errors in thousands of documents is not very easy.

In ABBYY FlexiCapture, there are several ways to locate markup errors.

First, you can train the classifier and classify the training set. This is not very useful for finding out the quality of real-life classification, but it is an excellent way to identify markup errors or technology limitations. You should aim for a quality of 98% or higher.

Once you have classified the test set, you can use tools like the confusion matrix that will help you focus your attention on the biggest issues and then drill down to specific errors. When you see a classification error in the test set, e.g. an invoice has been wrongly classified as purchase order, there may be two reasons for that:

1. The training set may contain a markup error, e.g. some invoices may have been wrongly assigned the class of purchase order. In this case, you need to find this error in the training set and correct it.
2. There are no similar-looking invoices for a particular invoice in the training set, whereas some purchase orders in the training set look very much like this unique invoice. In this case, you need to add some invoices resembling this unique invoice to make it more representative.

Use **Show Similar Pages from Result Class** to investigate the first possibility, and **Show Similar Pages from Reference Class** to investigate the second. The result will be a list of pages sorted by similarity to the document that has caused the classification error. Use the **Show Similar Pages** command to find similar pages in the entire batch, regardless of their reference classes.

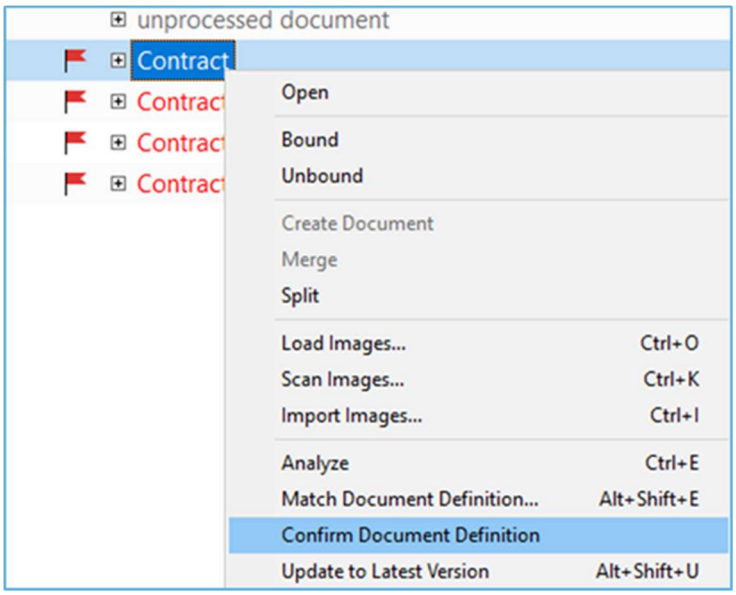
Search for similar pages relies entirely on image features and does not employ text analysis. For this reason, this feature will best work on documents with distinctive, visible differences.

5.2.7. Verification of classification results based on confidence

Classification can be used for document type identification and document assembly. Classification is done at page level, so if the classifier is not completely sure about the class of a page, verification is required.

Pages and documents that require verification are highlighted in red. If a document has been identified and assembled correctly, the operator can confirm the Document Definition at document level manually by clicking the **Confirm Document Definition**

command on the shortcut menu. Otherwise, the operator can change the Document Definition by clicking the **Match Document Definition** command or by using a new panel specifically designed for quick assignment of Document Definitions (click **View > Side Panels > Set Document Definition** to show or hide the panel). If a document has been assembled incorrectly, the operator can use drag-and-drop and special commands to put pages in their appropriate documents.



Verification of classification results can be part of the verification stage, or it can be a separate document assembly review stage.

In a classification script, you can specify thresholds indicating when verification is required. By default, this property is specified for each page automatically by the classifier, but you can also use a script to estimate the probability of the best class and the distance to the second-best hypothesis and decide if verification is required for this page. For example, if there is a 90% probability that a page is an invoice, then probably no verification is required, but if it is also an 80% probability that it is a credit note, then you may want the page to be verified. If one of the pages in a document requires verification, then the entire document will require verification.

Sometimes, the mere fact that a matching Document Definition has been found for a document is sufficient evidence that the class of the document has been was detected correctly and there is no need for additional verification. For such Document Definition sections, the **Automatically confirm section type when matched** option should be selected.

5.3. Recognition

5.3.1. Improved Arabic OCR

	Speed (pages per minute)	**Accuracy**
Thorough	16.8	79.59% - words, 93.6% - characters
Fast	45.79	73.7% - words, 91.62% - characters

Fewer characters require verification after recognition.

5.3.2. Improved Japanese OCR

Japanese (Modern)

Japanese (Modern) can be used for contemporary business documents because besides Japanese characters it also includes the English characters and the four most commonly used Greek characters.

	Speed (pages per minute)	**Accuracy (in characters)**
Thorough	37.7	94.25%
Fast	40.6	96.57%

Japanese

Traditional Japanese language.

	Speed (pages per minute)	Accuracy (in characters)
Thorough	35.45	96.75%
Fast	60.18	96.58%

5.3.3. Improved Korean OCR

Korean

	Speed (pages per minute)	Accuracy (in characters)
Thorough	33.53	96.43%
Fast	37.63	95.60%

5.3.4. Field recognition improved for Japanese documents

Some Japanese texts should be read left to right and some should be read top to bottom. Text direction is detected automatically. Still, if only one direction is possible in a particular case, this can be indicated in the **Direction of CJK text** option in the field properties.

5.3.5. Patterns (.fbt) can be applied to improve recognition of specific characters within fields*

Patterns can be used to improve recognition of specific characters. Patterns can be trained in ABBYY FineReader 14 or in ABBYY FineReader Server 14 and then applied to ABBYY FlexiCapture projects.

To load a pattern into ABBYY FlexiCapture, open the recognition properties of a text field, and in the **Advanced** section click **Modify**… > **Use pattern** and select an *.*fbt* file.

Note: Patterns for CJK languages cannot be trained.

5.3.6. The language of text in fields is available in scripts to enable integration of third-party translation services

The language of text in fields is automatically detected during recognition and then is available in scripts via the *Ifield.GetLangs()* method. This is useful if you need to send this text to a third-party translation service.

Note that when languages are detected during recognition, only languages allowed for field recognition will be considered as candidates. If the operator types the value of the field manually, this property will always return to the default system language.

5.4. Detection of page orientation will ignore manually rotated pages

When applying Document Definitions and recognizing documents, the program will not detect the orientation of pages whose orientation has been manually corrected by the user.

5.5. Export

5.5.1. Export to Managed Metadata column type in SharePoint

Data can now be exported to the Managed Metadata column type in SharePoint.

5.5.2. Export to XLSX

Field data captured by ABBYY FlexiCapture can now be exported to *.xlsx.

Bug fixes

Issue Description
Machine learning on user side for line items. After training, the region of the column sometimes is expanded and values from neighboring columns are captured.
Machine learning on user side for line items. After training, the region of the 1st line sometimes is expanded to half the page.

Issue Description
An error occurred when working with document sets in a specific scenario: after the 1st document was recognized, it was immediately opened at the same moment as the next document was added to the same document set. Opening the second document sometimes resulted in an error.
Instances of a repeating group were cut to the size of the 1st instance during export to the database.
The "Type mismatch" error occurred when exporting an empty value to a column of type *int* in an SQL database table.
In the Document Definition editor, if Japanese (Modern) was selected alongside with any other language in the text field properties, it was changed to "Japanese."
DPI did not change when change of DPI was included into an image enhancement profile.
An error occurred when exporting to PDF-S using *SaveAsStream().*
ABBYY FlexiCapture for Invoices. Values less than 0.1 were presented as .e-002 if a data set was populated with values from the database.
In a project where index fields with regions were used, the first change made by the verification operator was ignored and only the second and subsequent changes were correctly applied.
Verification Station. We no longer expand a document in the details list when it is opened by the verification. (The behavior of version 11 has been restored).
An error occurred if a line in a document was set as the identifier by clicking the **Hint** button in the identifier properties.
When using *IdrawContext,* some UI elements were incorrectly displayed.
Recognition results could differ slightly in ABBYY FlexiCapture and ABBYY FlexiLayout Studio even if the recognition settings were identical.
An error occurred processing a PDF file with a font which had an empty OS/2 table.
The field of type “Picture” on a fixed form did not change the value on the data form when the verification operator modified its region.
ABBYY FlexiLayout Studio. Adding a character to the alphabet of an element of type CharacterString could result in an error.
An error occurred when the user matched a Document Definition manually and enabled recognition afterwards.
An error occurred when uploading a project to the server with incorrect rules (sometimes incorrect rules were part of old version of Document Definitions). Now old versions of Document Definitions are ignored when uploading projects and rules are be checked more carefully (new checks have been added).
Export to PDF/A always created a text layer. Now it is possible to switch off the **Create searchable PDF** option and create a PDF/A document without a text layer.
Using the method *AnalyzePageObjects()* method in a script could result in an error.
Incorrect or empty text was returned by classification scripts that used *Page.FullText.*
EAN 13 separator barcode was not detected on some documents.
A “Paragraph” element created in ABBYY FlexiLayout Studio failed to capture all lines between reference elements.
Web Stations. In rare cases, line items on the data form on the Web Verification Station were replaced or some of them disappeared when the user resized browser window.

Issue Description
Web Stations. The external links to the "Hardware and software requirements" and "Installation instructions" sections on the personal web page were incorrect.
Web Stations. On the data form on the Web Verification Station, manually added elements of a repeating group were not displayed if the **Show as a table** option was selected.
Web Stations. The Japanese GUI of the Monitoring Station linked to the English version of Web Help.
Web Stations. In rare cases, after checking rules and rerecognition, the Web Verification Station loading screen froze.
Web Stations. No project was available on the Web Verification Station after installing ABBYY FlexiCapture on the server for the very first time.
Web Stations. On the Web Verification Station, the Invoice Number field could not be recognized correctly on some invoices.
A dash, when placed alone on a line, was sometimes not recognized.
When the PDF text layer was used to help OCR on some CJK documents, the OCR result could be significantly worse compared to using just OCR alone.
Sometimes the first attempt to edit a document during verification was ignored. The error was caused by the LastEditor service field, which is updated each time the document is opened and causes the document statistics to be recalculated.
An error occurred when processing pages with anisotropic resolutions. Starting from this Release 3, resolution will be adjusted for such pages prior to processing.
Filtering by stage has been removed for classifier training batches and field extraction training batches, as it is unnecessary.
When applying a Document Definition manually, text objects were not rotated properly on pages whose orientation was corrected.
In some cases, two lines of Japanese text were merged into one.
The Processing Server crashed when attempting to send a very large message.
Two-line static text was not found if the regions of the lines overlapped vertically. Now when you can create FlexiLayouts in the advanced code window in FlexiLayout Studio, you can specify allowed percentage of overlap for StaticText elements. For example: AllowIntersectPercent(15) For existing projects, we recommend using the following registry key to specify allowed percentage of overlap: Computer\HKEY_CURRENT_USER\Software\ABBYY\FlexiCapture\12.0\DAForms\geometryPhraseIntersectPercent Computer\HKEY_CURRENT_USER\Software\ABBYY\FlexiLayoutStudio\12.0\DAForms\geometryPhraseIntersectPercent
Digits could be incorrectly recognized as letters in fields with monetary amounts. In the “US Invoice” project, a regular expression has been added to improve the quality of recognition. If the regular expression does not appear after you upgrade your “US Invoice” project, copy it manually from the updated “US Invoice” project provided with Release 3.
Unable to create or modify records in a dataset via SOAP API when specifying values for numeric columns.
When text typed on the Web Verification Station did not fit into the text box and new lines were expected, sometimes no new lines were added and the text was displayed outside the text box.
Operators of Web Verification Stations could not get tasks when working via Fortinet SSL VPN.

Issue Description
Sample custom reports in Russian were visible in installations in other languages. Now the Russian sample reports are only visible in Russian installations. For all other languages, only English samples are shown.
Users could not change the date filter for a report without refreshing the web page if an empty period had been selected when creating the report on the Monitoring Station.
On the Monitoring Station, a wrong stage name was displayed for a custom stage in the **Final Stage** column of the event log.
No projects were available on Web Verification Stations if ABBYY FlexiCapture was first installed on a machine without ASP.NET.
On some rare occasions, verification tasks could not be opened on Web Verification Stations using Windows authentication.
When using the Web Scanning Station, temporary files were not automatically deleted from C:\\Users*UserName*\AppData\Local\Temp.
After upgrading ABBYY FlexiCapture, an error message saying "Could not connect to the database server" was shown on the Login Page in some browsers (e.g. in Internet Explorer and Firefox).
In some cases, when a user opened the "Send to Stage..." window on the Web Verification Station, the system batch name appeared as a stage name prefix.

Known issues

Issue Description
NLP text segments are not supported on the Web Verification Station.
Manual entry of line items on the Web Verification Station will be optimized in upcoming updates.
Image enhancement profiles are not supported on the Web Verification Station and on the Scanning Station. Image enhancement profiles created on the Project Setup Station or on the Scanning Station can be applied automatically during processing, but profiles cannot be created or applied on web stations. Only simple operations with images are available.
Document sets are not supported on the Web Verification Station and on the Web Scanning Station.

