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 and updates.

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.

Deep learning for invoice processing

This paragraph is updated and correspond to the Release 3 Update 1.

A new technology based on neural networks is now used to extract header fields and line items from invoices.

Network is trained to extract the following header fields:

- Invoice Number
- Invoice Date
- Total
- PO number
- Vendor Name
- BU name

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+%**.
- Purchase **Order Number** can now be extracted out-of-the-box without database lookup or regular expression with quality 92,4% and higher. See also “ABBYY FlexiCapture for Invoices: Purchase Order Number field extraction is improved”.
- **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”).

Neural Network is trained to extract Header fields and LineItems on 25000 invoices. Invoices from the countries listed below have been used to train ABBYY FlexiCapture.

- Austria,
- Australia,
- Belgium,
- Bulgaria,
- Canada,
- Czech Republic,
- France,
- Finland,
- Germany,
- the UK,
- Greece,
- Japan,
- Hungary,
- Ireland,
- Italy,
- Lithuania,
- Malaysia,
- Netherlands,
- New Zealand,
- Poland,
- Singapore,
- Slovakia,
- Spain,
- Sweden,
- Switzerland
- 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. For this benchmarking we have used almost 10K invoices with rather good quality of source images from main countries. Quality on scans with low resolution and photos can be lower. The field is counted as correctly detected when its region is correct (coincides with reference region).

Country	Invoice Number	Invoice Date	Total	Line Items	
				Rows	Quantity, Unit Price, Total
AU	92.7	97.3	86.5	79.7	65.0
CA	93.6	95.3	90.7	83.7	56.1
DE	92.1	95.0	94.4	70.8	58.5
ES	99.0	97.3	95.5	89.6	60.0
FR	93.2	92.3	85.9	76.0	60.4
GB	95.0	95.2	92.4	77.4	62.3
US	91.4	93.0	85.8	85.7	72.8
JP	77.4	93.2	71.3	69.3	54.4
Average	91.8	94.8	87.8	79.0	61.2

Notes:

- As the new technology requires more computing resources, we recommend installing the x64 version of the program and having at least 1 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.

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 1 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, Hungarian and Italian.

Notes:

- The web stations are available in all of the above listed languages except Serbian. The web stations are also translated into Chinese Traditional.
- 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 Update 2 Patch 1	**1299/43**	**12.0.3.4040**	**176094**	**16.1.1014.15**	**2020.08.06**
Release 3 Update 2	1299/42	12.0.3.4038	171689	16.1.1014.14	2020.07.23
Release 3 Update 1 Patch 3	1299/41	12.0.3.2655	129595	16.1.815.2	2020.04.24
Release 3 Update 1	1299/39	12.0.3.2634	108384	16.1.814.24	2020.03.06
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

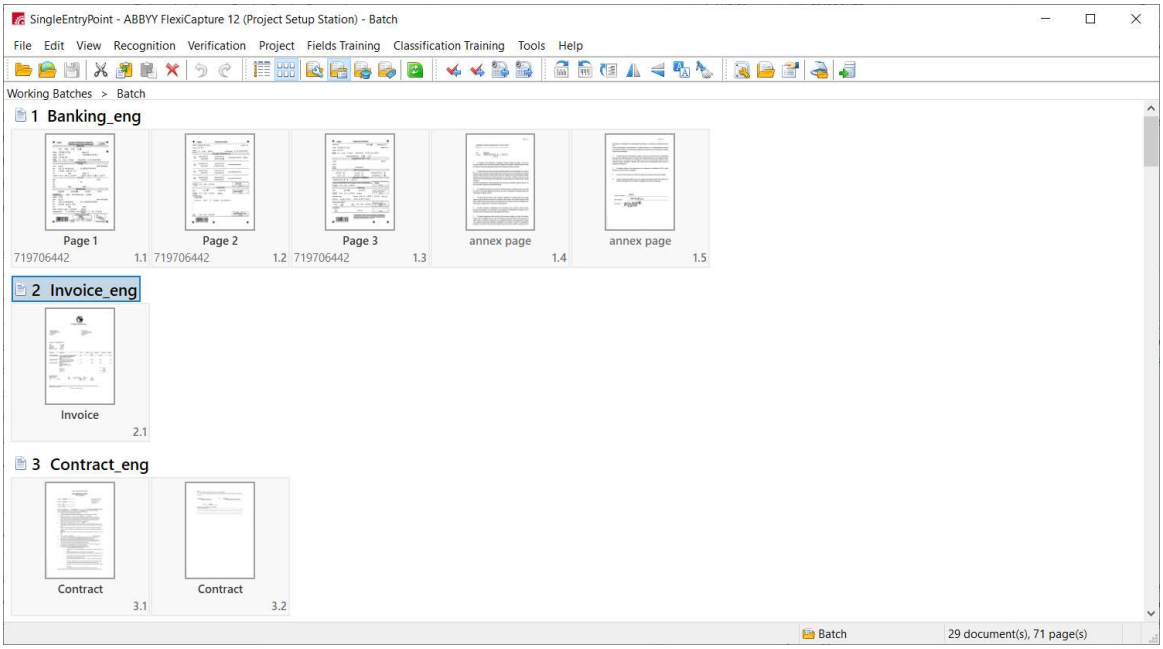
When one physical set of pages represents several documents from a business point of view, it may be necessary to process such documents using different validation rules or send them to different approval workflows. In this case an operator can create a duplicate of an existing document and manually 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

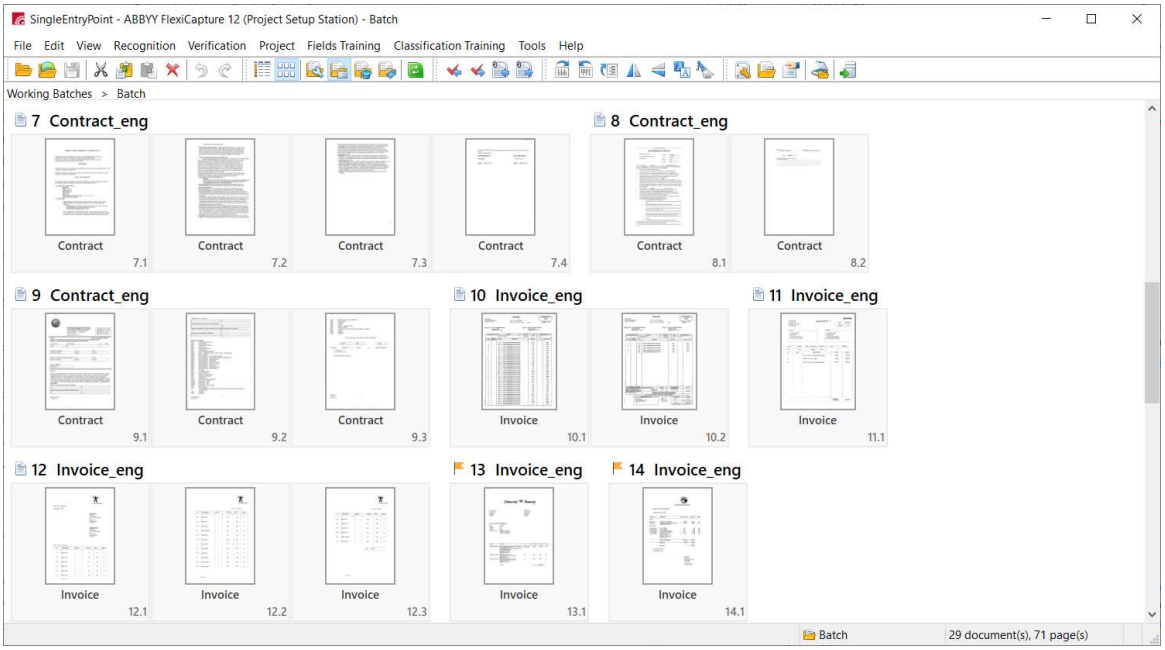
If you work with batches consisting of multiple one-page documents, you may want these documents to take up all the screen space, so that you can see as many documents as possible at once. On the other hand, if you work with multi-page documents, to you may want to start each document on a new line to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

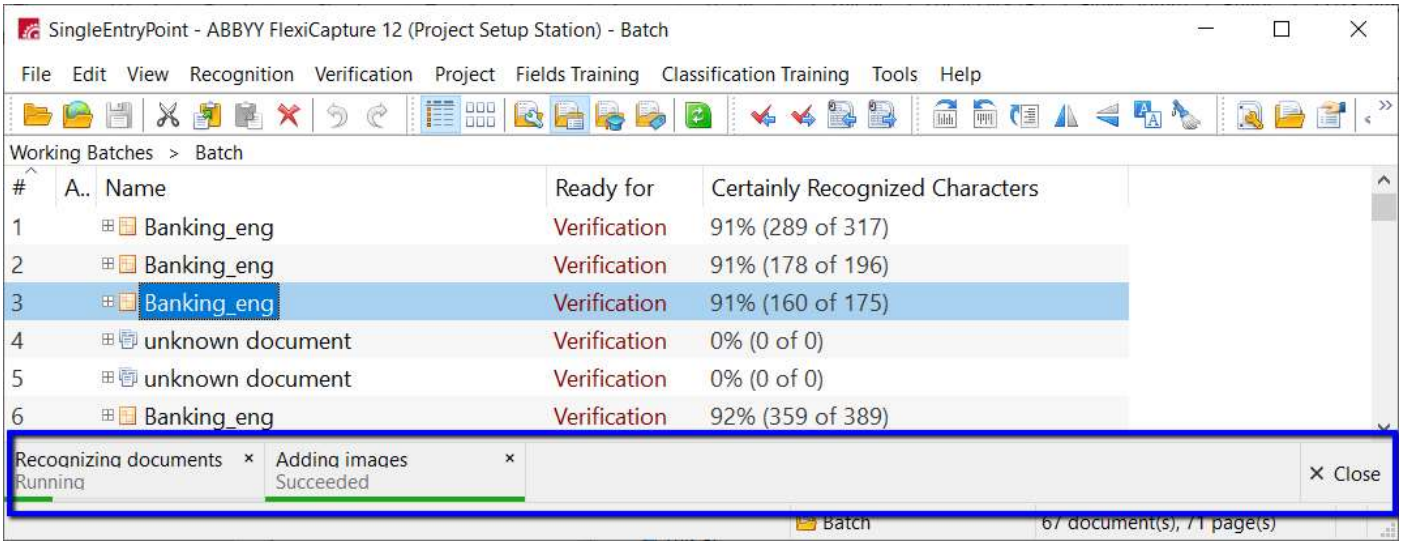


Documents use up all screen space:



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. Limitations of Using Internet Explorer 11 browser for Web stations

Due to the known limitations of Internet Explorer 11, you may observe the issues while processing dozens of pages at once or having numerous verification tasks with ABBYY FlexiCapture. Using Internet Explorer 11 browser for an extended period of time can cause the browser to consume a lot of the machine's memory and slow down the client performance and eventually crash the browser. For more details see “Limitations of usage of Internet Explorer browser for Web stations”.

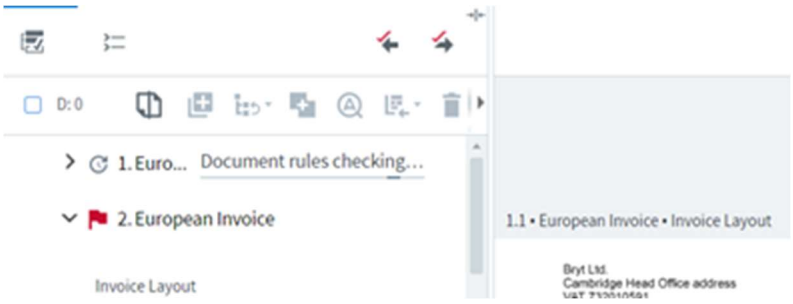
2.2. Web verification improvements

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

2.2.1. 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.



2.2.2. 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.

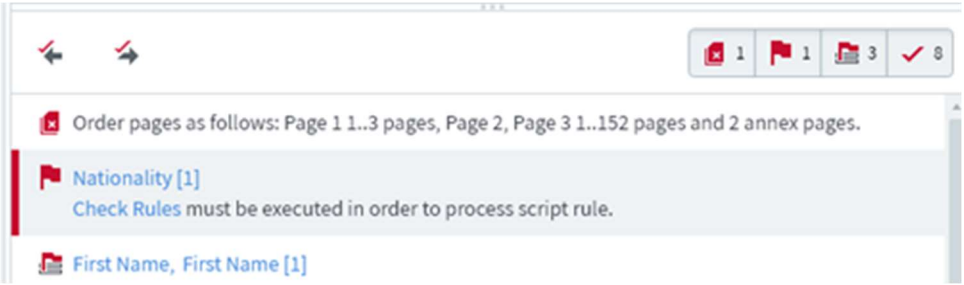
For your script to be executed without calling rules, the objects used in the script should not require rules to be checked on Web Stations. If your script contains at least one object that requires rules to be checked on a Web Station, the "Check Rules" button will be used for checking rules. Detailed 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."

2.2.3. 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.

2.2.4. 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

2.3. Invoice processing improvements

2.3.1. 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.

2.3.2. 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.

2.3.3. 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.

2.4. Other improvements

2.4.1. 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.

2.4.2. 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.

2.4.3. 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.

2.5. 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.

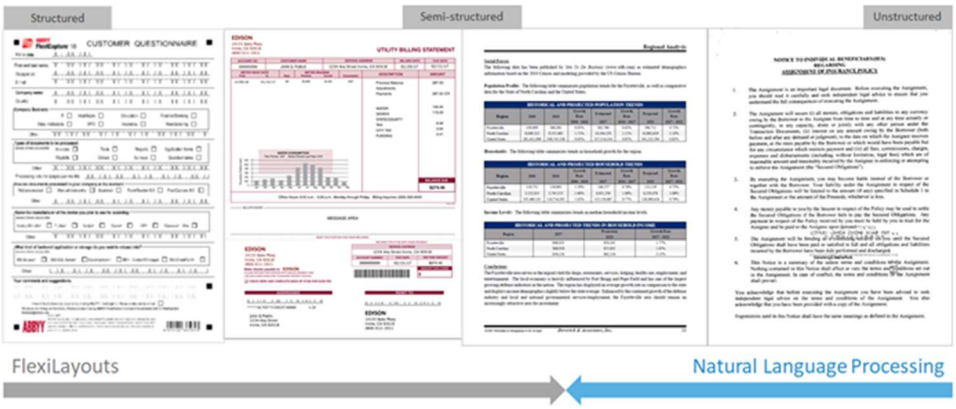
2.6. Ability to download exported results on the Web Verification Station

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.

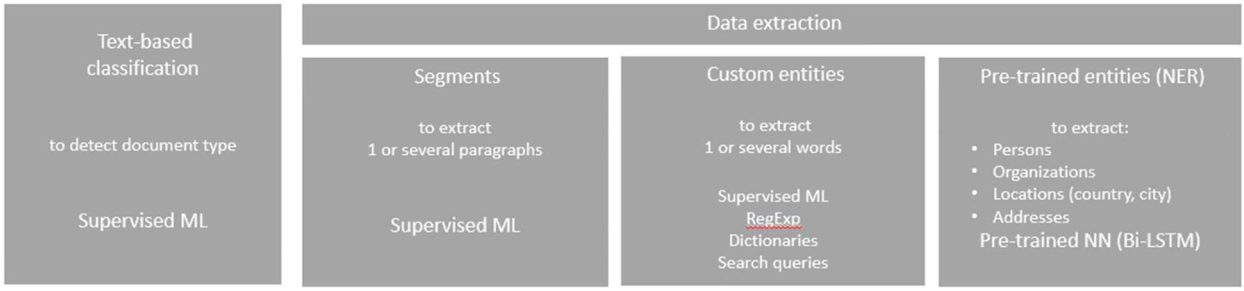
Semi-structured documents. If you examine a few invoices from different vendors, the data will seem to be positioned randomly on the page. On the other hand, if you examine several invoices from the same vendor, you will notice that there is some structure. If you examine a sufficiently large number 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

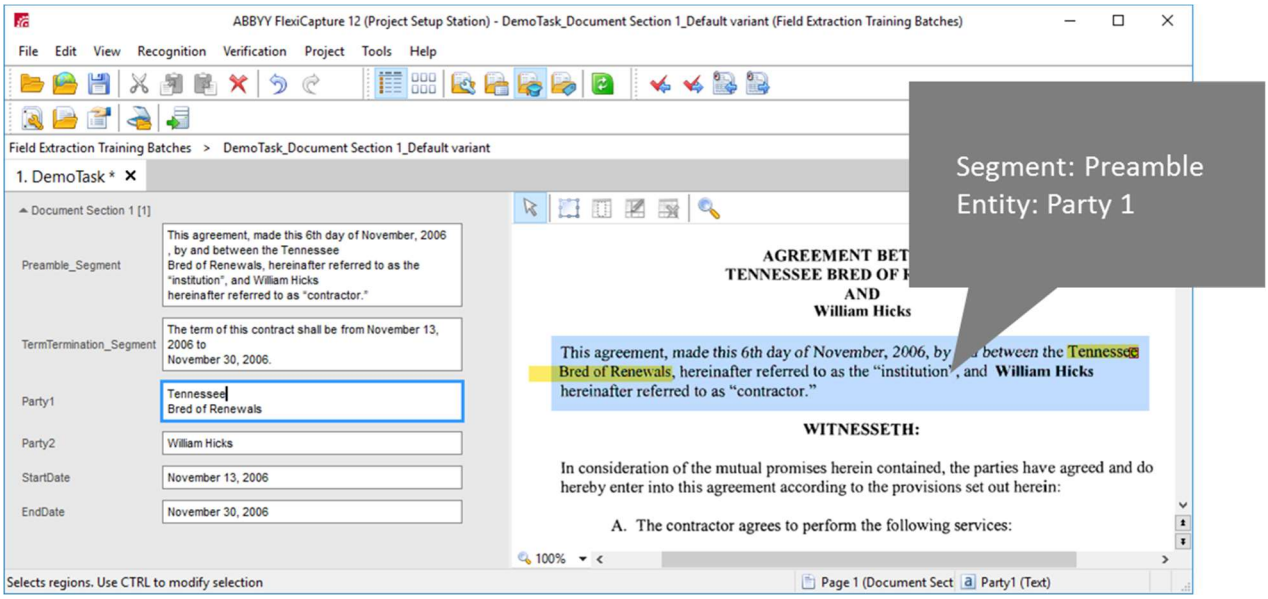
To extract data from a document, you typically first need to identify the type of your document. ABBYY FlexiCapture contains advanced classification technology that uses both image and text based[1] features to identify document class. Document classification enables the program to determine which Document 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.

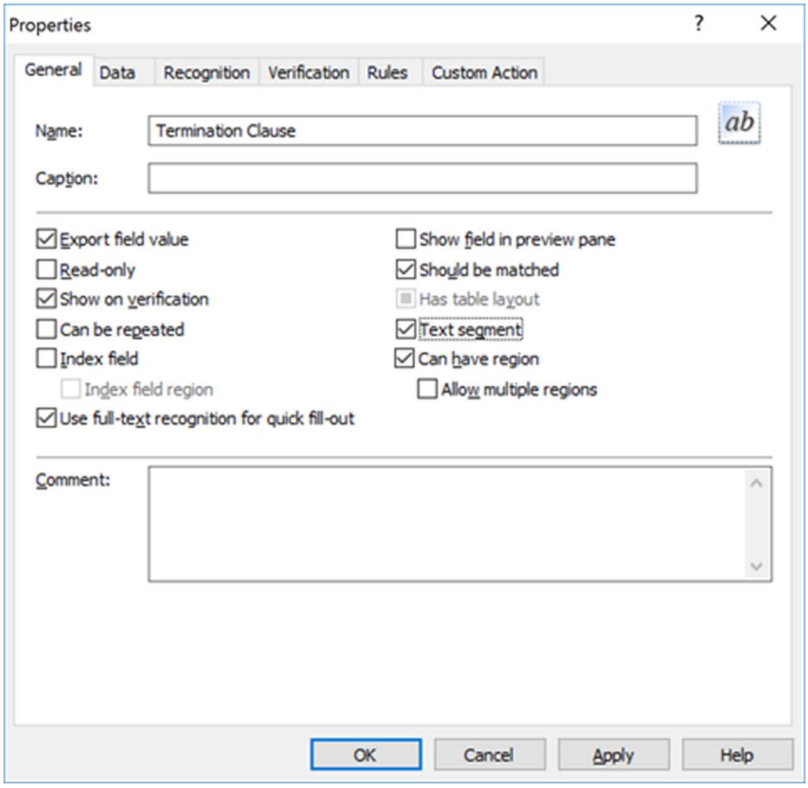


3.3. How to train NLP Machine Learning model?

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.

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



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.

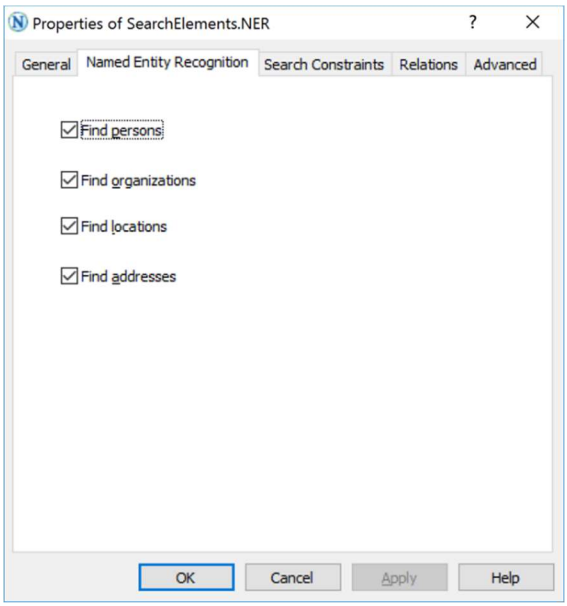
3.4. What if you do not have enough documents for training? NER (Named Entity Recognition), RegExp (regular expressions), dictionaries, search queries.

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 FlexiCapture 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. 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.

4.2. 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.

4.3. 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.

4.4. 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

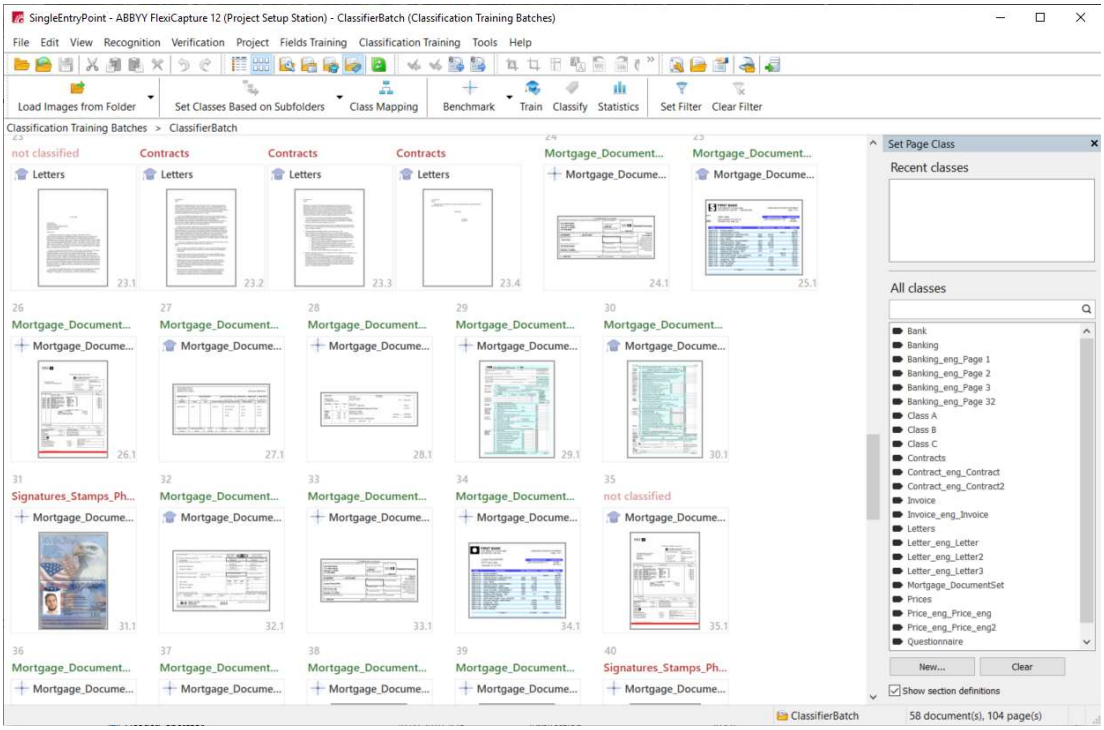
New methods have been added to *IeditablePictureObject,* which allow you to apply any image enhancement profile from within a script.

- **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:

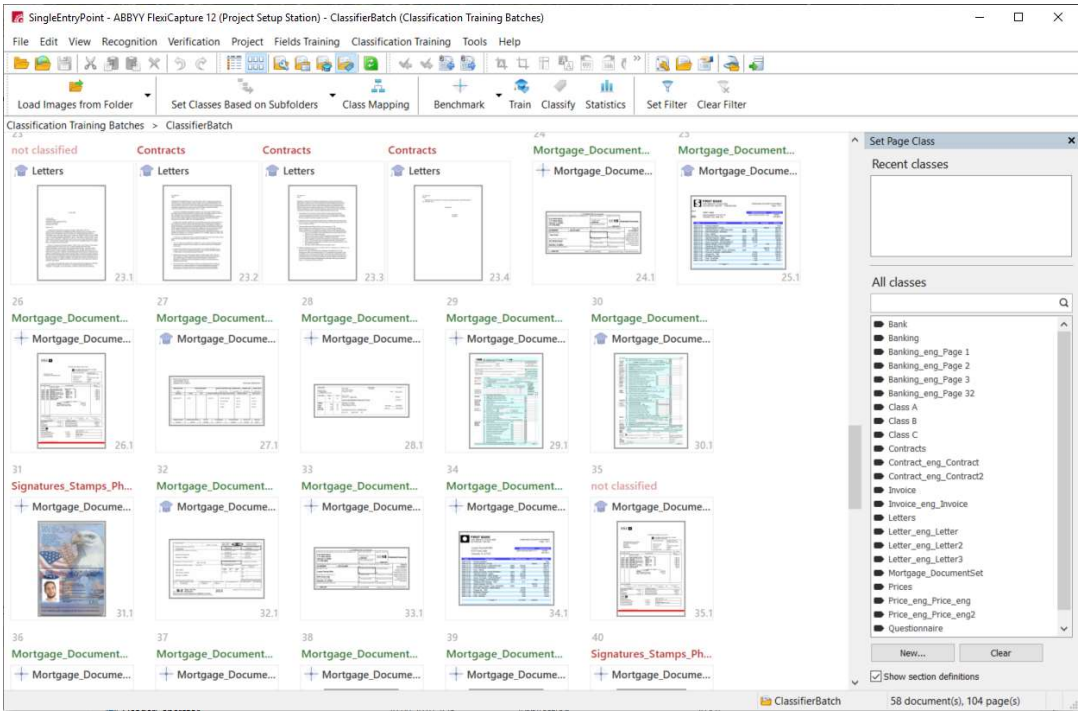
1. Import documents and assign reference classes to them.
2. Train the classifier and analyze any issues.
3. Improve the training set accordingly.
4. Establish correspondences between classes, Document Definitions, section variants, and page roles (i.e. the first page, any page).
5. Publish the classifier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

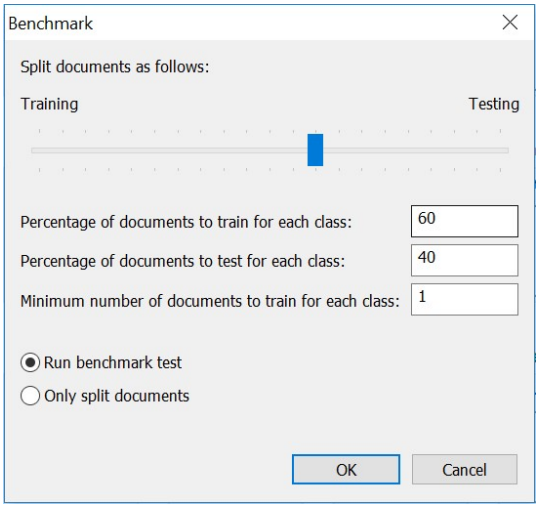
5.2.3. Thumbnail view optimized for handling classification batches; a special panel added for setting page classes quickly

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

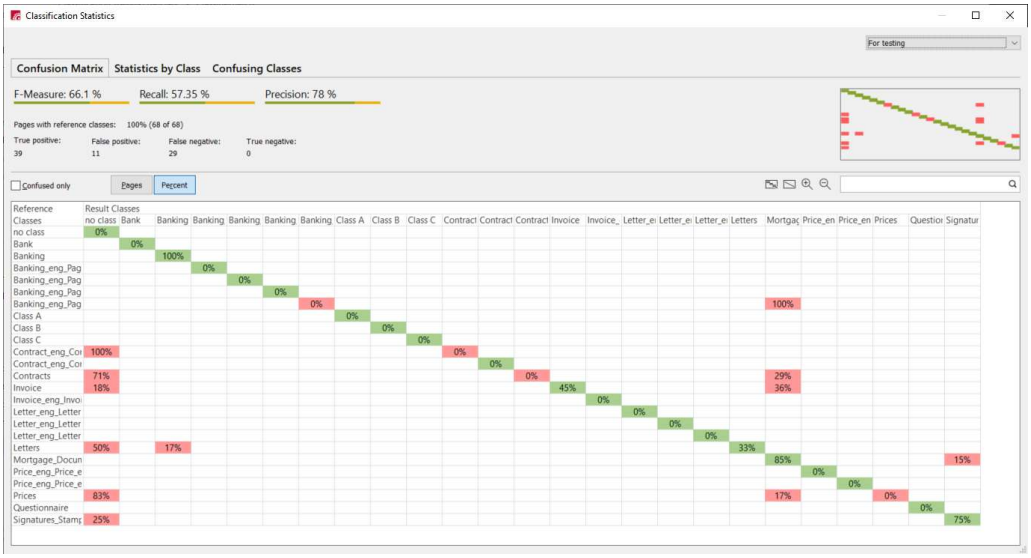
When the benchmark test is finished, statistics are shown for the test set. In addition to the main quality indicators, such as F-measure, Recall and Precision, other statistics are presented on the following three tabs:

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.

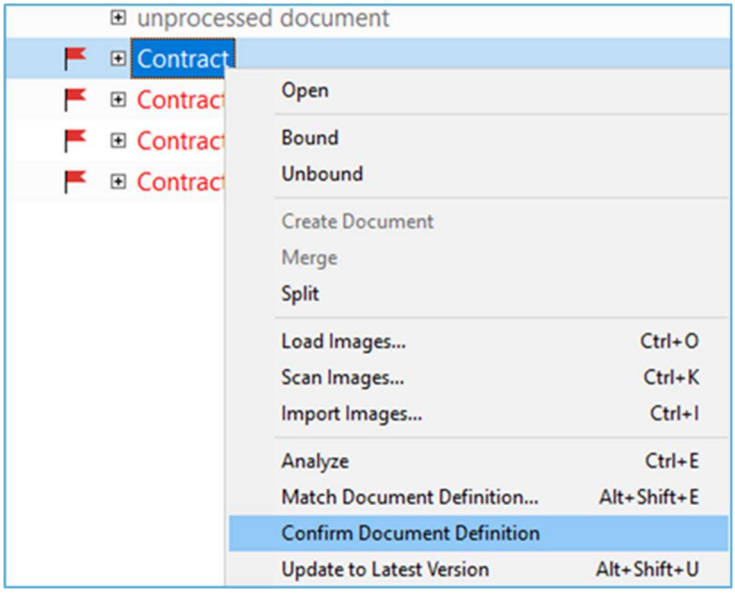
Use **Show Similar Pages from Result Class** to investigate the first possibility, and **Show Similar Pages from Reference Class** to investigate the second. The result will be a list of pages sorted by similarity to the document that has caused the classification error. Use the **Show Similar Pages** 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.
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.

Issue Description
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.
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:

Issue Description
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.
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.

Update 1

Technical information

Release	Part #	Build #	Installer Build #	OCRT build #	Release date
Release 3 Update 1	1299/39	12.0.3.2634	108384	16.1.814.24	2020.03.06

New functionality

1. Import from an SFTP server

An SFTP server can be used to send documents to the ABBYY FlexiCapture-based cloud service over a secure connection. It can be also used with a common FlexiCapture installation. Documents are imported into ABBYY FlexiCapture automatically, making all the Hot Folder capabilities available to the ABBYY FlexiCapture-based cloud service.

To use the SFTP import feature, follow these steps:

1. Select **Hot Folder** for the import source.
2. Select **SFTP server** for the Hot Folder type. Specify the address of your SFTP server. If you are using a non-standard port, specify it after the server address (e.g. sftp://address:port/).
3. Click the **Settings...** button and enter your login and password in the **Network Settings** dialog box.
4. Make sure the following permissions are given to the SFTP server:
    a. File permissions: Download, Upload, and Delete
    b. Directory permissions: List, Create, and Delete

Note: Only password-based authentication is allowed.

2. Export to JSON

Starting from this release, JSON can be selected as a file saving format when exporting data to a file.
You can:

- Specify additional data to be exported (e.g. field locations, errors, or recognition quality data).
- Select the **Preserve value type on export** option to save an XML schema with field type descriptions.
- Select an encoding (see the “Export file formats” section).
- Specify field names, if necessary. To do this, click **Field Mapping** and, in the **Field Mapping** dialog box that opens, specify the field names.

Known limitation: Decimal values will be enclosed in quotation marks (e.g. "14.25").

3. ABBYY FlexiCapture for Invoices: Purchase Order Number field extraction is improved

In ABBYY FlexiCapture 12 Release 3 and earlier, Purchase Order Numbers could only be extracted using databases or regular expressions. Now PO numbers are extracted using neural networks and such extraction is enabled by default.

Extraction based on neural networks is supplementary to other methods so it will work in case PO number was not extracted with other methods such as database lookup or search using regular expression.

The quality of extraction out-of-the-box was measured on testing batches with normal quality of images for OCR:

Country	F measure	Recall	Precision
AU	90,25%	98,43%	83,33%
CA	94,85%	99,63%	90,51%
DE	90,85%	98,69%	84,17%
US	93,94%	98,94%	89,42%

If for some reason you need to disable Purchase Order Number extraction based on neural networks add the register key and set a value as described:

[HKEY_CURRENT_USER\Software\ABBYY\FlexiCapture\12.0\DAForms]

EnableNeuralNetPONumber = false

4. ABBYY FlexiCapture for Invoices: extraction of BU name and Vendor name based on neural networks

Neural Network is also trained to extract Vendor name and BU name. This extraction will work in both cases: no vendor or BU databases are connected, or databases are connected but database lookup was not successful.

5. ABBYY FlexiCapture for Invoices: Purchase Order number extraction trained on the user's side

Extraction of Purchase Order Numbers can now be trained on the user's side. For several Purchase Order Numbers in one document, only the first instance will be used for training.

In fact, another non-repeatable field is trained because of the restriction to train repeatable group. Then the value is automatically copied back to the instance of the group. That is why for new invoice projects, training of PO number will be enabled by default but when using an invoice project created in an earlier version, additional adjustments should be done. After project update to the new version, please follow these steps:

1. Open your project in ABBYY FlexiCapture 12 Release 3 Update 1. Your project will be updated.
2. Open the **Document Definition Properties** dialog box, click the **Invoice Settings** tab, and make sure that the following options are selected:
3. Purchase order matching
4. Thorough extraction of invoice header fields
5. On the **Event Handler** tab, select **After document state changed**.
6. For the **TrainablePO** group of fields, select the **Copy PO data to TrainablePO** rule.
7. Check the recognition setting of fields from **TrainablePO** group and if they have **Do not recognize** setting on Recognition tab, change it to ***Standard Recognition***.
8. Start training PO number extraction.

6. ABBYY FlexiCapture for Invoices: improved detection of line items

In case of multi-line product or service descriptions, a line item could be incorrectly separated into several lines. Extraction of such line items has been improved so now the program interpret them like single line item with long description. Please note that the extraction of multi-line description of the last LineItem on the page may fail.

If you need to extract a lot of multi-line description, the mode of extraction of LineItems can be changed using registry key:

[Computer\HKEY_CURRENT_USER\Software\ABBYY\FlexiCapture\12.0\DAForms]

"EnableExtendingDescriptionInLineItems" = true

7. Updated EULA

The EULA has been updated. Besides English, the EULA is now available in Russian, Japanese, German, French, and Spanish.

8. Web Verification UI translated into Traditional Chinese

The UI of the Web Verification, Web Scanning and Web Capture stations is now available in Traditional Chinese.

9. Help of Scanning Station is translated into German and French

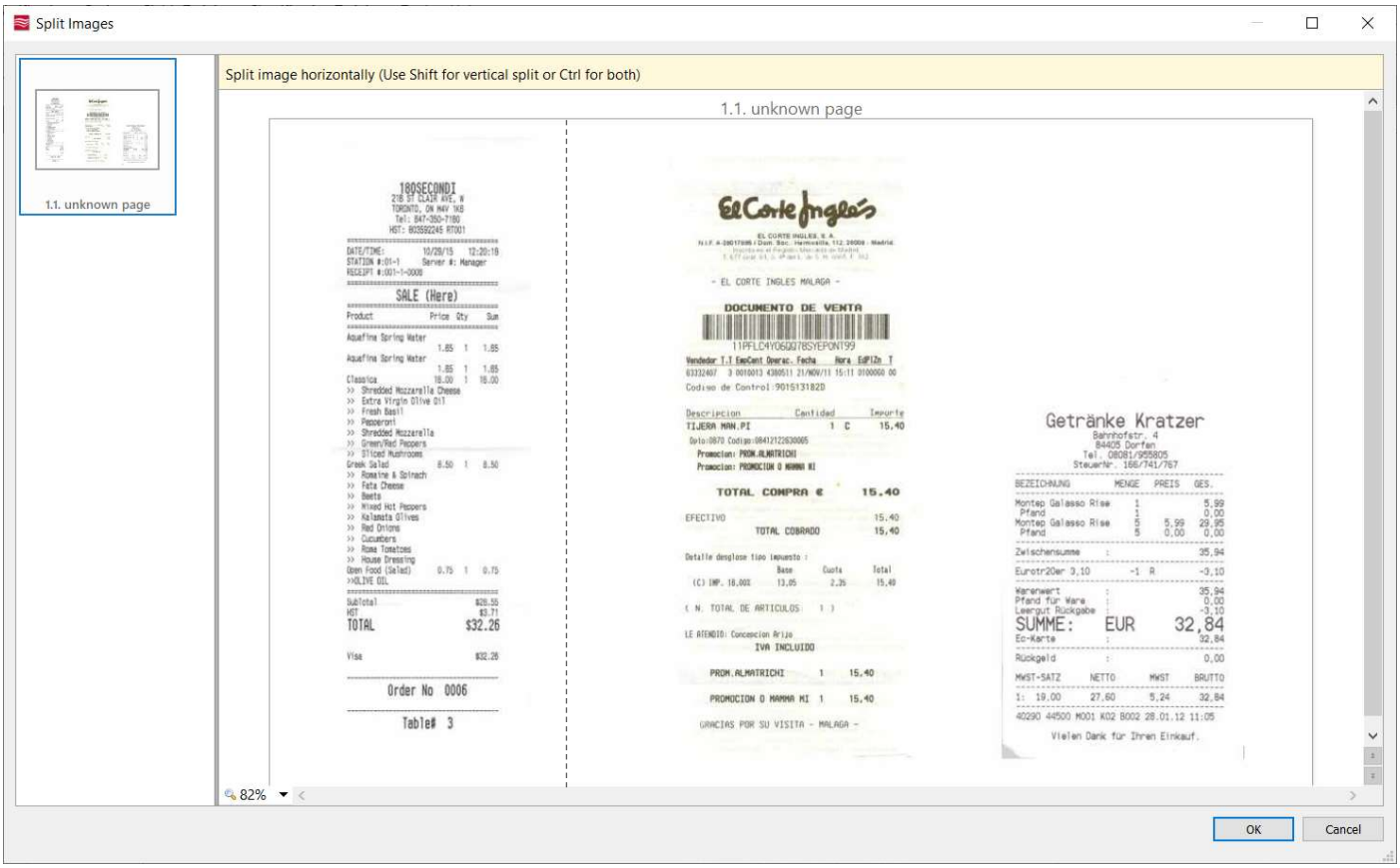
Help file for Scanning Station is available now on German and French.

Online help in French
Online help in German

10. Ability to crop and split images during verification

When multiple documents were scanned together (e.g. when scanning IDs or receipts), it was only possible to split the resulting image on the Scanning Station. Autocrop was only available on the Scanning Station or via an image enhancement profile. Now Automatic crop, Split and manual Crop tools are also available to verification operators. They can use corresponding buttons on the toolbar:

Splitting receipts scanned on the same page:



11. Ability to delete a project from the list of recent projects

Sometimes there is no need to keep a project on the list of recently opened projects. For example, the administrator may upload a project to the server and have no need to work with its local version. Now such redundant projects can be deleted from the list of recently opened projects so as not to confuse the operators.

Note. The feature is only available on Windows clients.

12. Ability to export classification statistics to a CSV file

Classification statistics can now be exported to a CSV file for further analysis.

To export classification statistics to CSV, open the **Classification Statistics** dialog box, click the **Export Statistics…** button, specify a path and file name, and select the type of statistics to export. The following export options are available: **Summary**, **Statistics by class**, **Confusing classes**, **All classes**.

The exported statistics will correspond to the set of documents selected in the **Document type** field: **For testing**, **For training** or both.

13. ABBYY FlexiCapture Web Verification Station improvements

1. Performance improvements for large data forms:
   - Smooth scrolling for data forms containing over 600 line items

- The GUI is now more responsive when correcting large numbers of errors

2. Working with links to existing fields and with the summary section is now supported. For more information, see these help articles: Link to existing field, Summary section.
   Please note that you can currently work with links to existing fields and with the summary section only within one document (e.g. a large document with multiple sections). The full functionality for working with document sets is not currently available on the web stations.
3. Field values of the following types can now be normalized: int, currency, date, and time.
4. Opening tasks sped up on Web
   The Web Verification station has been redesigned. Operator waiting times for starting verification tasks were reduced by half.

14. ABBYY FlexiCapture Developer's Package

Release 3 includes an ABBYY FlexiCapture Developer's Package for use with ABBYY FlexiCapture Cloud and the ABBYY FlexiCapture SDK. The Developer's Package can be downloaded and installed separately and contains only the following components required to develop and maintain FlexiCapture projects: Project Setup Station, FlexiLayout Studio, FCAdminTools, License Manager, and Licensing Server. To develop projects with NLP features, NLP components must be installed.

To use Developer's Package customers will need license for FlexiCapture Distributed with a permission to use Project Setup Station.

Bug Fixes

Issue description
A new database could not be created and an existing database could not be upgraded in Oracle 12.0.1.* due to a compatibility issue.
A classifier training task failed if the number of images in its training batch reached a certain limit.
Checking rules on web stations failed if working with WinAuth.
Multiple "-" symbols were not allowed in a login.
If a group contained only subgroups, syncing with AD removed all users from that subgroup.
On some scanner models, an error occurred if ISIS drivers were used for scanning.
Fixed forms became invalid after updating the project.
FlexiCapture for Invoices. Keywords from all recognition languages influenced field extraction for invoices in unrelated languages. For example, Russian keywords could hamper the processing of Swiss invoice. After the fix, if the country is detected, only languages related to that country will affect field extraction. If the country is not detected, the behavior will remain the same as before the fix.
While working on the web, a script rule failed if it used a shared assembly.
An error occurred when recognizing certain documents.
Japanese UI improved based on feedback.
Table column was not found if it contained "-1" values.
Project size reduced for projects with a lot of fields in Document Definitions (saving field settings in a workflow has been optimized).
The following security issues found during penetration testing have been fixed: • An operator with restricted permissions (e.g. a Scanning Operator) was able to close sessions of other operators. • An operator was able to postpone any task, even those already open and in use by another user.
Sometimes, the user needed to click repeatedly in order to select records in a data set.
In projects uploaded to the server, the **Disabled** parameter for training batches was ignored and trained layouts were still applied.
Excessive memory usage for Document Definitions with empty FlexiLayouts.

Issue description
After publishing a Document Definition on the server, an error icon was displayed even if there were no errors.
Occasional slow processing speeds when using a classifier or additional FlexiLayouts.
The Project Setup Station sometimes crashed when testing a Document Definition following a FlexiLayout update. The crash was caused by a specific location of data fields on the data form.
After updating a .NET assembly to a different version, it was necessary to recreate a reference to this assembly and save all the scripts once again.
During web verification, a document sent to back-end processing sometimes got stuck with an incorrect status.
An export task failed and would loop in some cases if the **Delete documents after export** option was on. Now this option will be ignored and a warning will be added into the log.
The orientation of some thumbnails was incorrect after rotating pages. The bug occurred only in local projects.
After SSO integration with IdP Auth, a ticket timestamp made the ticket invalid.
Automated processing failed when executing a script that edited a field with a long text string (1000+ characters).
An error saying "*Invalid character value for cast specification"* occurred when comparing a 5-digit number in a field of "Decimal" type.
Sometimes, an error occurred when a local project was used to update another project to a new version.
Field groups overlapped when scrolling data on a web station.
Sometimes, redundant indents occurred on the data form displayed on the Web Verification Station.
When using certain script rules, there were issues with loading the data form on the Web Verification Station.
In some projects, data could not be added to data sets on the Web Verification Station.
The user interface of the Web Capture Station became blocked when the user downloaded batch results locally.
Sometimes, after installing ABBYY FlexiCapture, the Administration and Monitoring Console would not open.
When the user specified an empty time period on the Monitoring & Reporting Station and then corrected the error, the interface still indicated the time period as incorrect.
In ABBYY FlexiCapture for Invoices, reduced RAM requirements for advanced field extraction technology involving deep learning. Previously, this feature required at least 1 GB of RAM per CPU core, now it is decreased and require 1GB. Peaks until 2.0GB can happen on complex multipage documents.
There were problems decoding e-mail bodies with unsupported characters when importing from mail boxes.
In FlexiLayout Studio, the order of columns in a column element could not be changed.
Fixed security vulnerability issue CVE-2019-20383. Users could elevate their rights to the Administrator or System level.
A program file required for recognition of Georgian texts was not installed. As a workaround, the required file could be copied manually onto the stations where recognition was carried out. Now necessary files are installed during application installation.
A training task failed if a batch contained documents for which several Document Definition versions were available.
Inline images and tables larger than A4 were not scaled correctly and so could not be processed.
Web Capture stations consumed too much backend resources, which led to random errors.
Applications crashed during data verification after latest update of Microsoft Windows.
ABBYY FlexiCapture Cloud
Sometimes, a task looped at automated processing stages.
The Project Setup Station could not open a tenant's project in the cloud if the tenant used its own license.
After upgrading the Application Server in a cluster, all the nodes had to be reconfigured manually in order to synchronize the values of the *machineKey* section in the *web.config* file.
After updating an expired license, the Processing Server had to be restarted in order to resume the processing.

Known Issues

Limitations of usage of Internet Explorer browser for Web stations

Due to the known limitations of Internet Explorer 11, you may observe the issues while processing dozens of pages at once or having numerous verification tasks with ABBYY FlexiCapture. Using Internet Explorer 11 browser for an extended period of time can cause the browser to consume a lot of the machine's memory and slow down the client performance and eventually crash the browser.

To avoid the unwanted interruptions of ABBYY FlexiCapture processing tasks ABBYY recommends you to use Microsoft Edge, Google Chrome or other browsers for the professional verification tasks. Nevertheless, ABBYY FlexiCapture web-stations are fully functioning in Internet Explorer 11, thus they should be used for ad-hoc processing preferably.

Important notes for users of Internet Explorer 11

1. In the event of a verification malfunction, the operator will not be able to resume the task immediately after the session is restored. This task will become available only after the inactive sessions have been purged. By default, the time since last user activity after which inactive session will be automatically closed, is equal to 120 minutes. You can change the setting *SessionInactiveLifetimeInMinutes* in the **web.config** file on the Application Server. We do not recommend purging inactive sessions too often. The recommended minimum setting is 10 minutes.
2. Starting from FlexiCapture R3 Update 1, each new task is opened in a new tab to optimize memory use when working in Internet Explorer 11. However, this also means that operators will not be able to receive new tasks automatically. To revert to the old behavior, open the **web.config** file for the web station and set the *IEVerificationNewWindow* key to *false*.
3. Purchase Orders matching does not work on Web Verification opened from Internet Explorer 11 in case of several instances of Purchase Orders Numbers instances on the same document. When clicking **Details** button, an error occurs (*ReferenceError: 'customActionContext' is undefined*). This error will be fixed in next updates.

Update 1 Patch 3

Technical information

Release	Part #	Build #	Installer Build #	OCRT build #	Release date
Release 3 Update 1 Patch 3	1299/41	12.0.3.2655	129595	16.1.815.2	2020.04.24

Bug Fixes

Issue description
Since Google Chrome (80.0.3987.116) and Firefox (73.0.1) when working in Web Station sessions weren't closed. It led to run out of licenses and "Cannot get license ticket" error message.
Results of verification were lost when leaving the last field on the form during verification on the Web station.
There was a vulnerability that allows a user from one tenant to get access to another using Administration and Monitoring console request.

Also included Bug fixes from Patch 1

Issue description
Rich clients silently closed if a huge table with LineItems was removed
Import in all tenants was blocked if a test copy of any project was created
Training failed for some mark-up
Default roles were sometimes deleted when editing project properties
The Web Verification Station didn't close the session when changing a project
Repeating fields were not reflected on Web stations

Known Issues

Limitations of usage of Mozilla Firefox browser for Web stations

When using Mozilla Firefox (version 72 and later), an active task may hang if you close the browser window while the task is still in progress. The task will be automatically returned to its appropriate queue once the inactive session is purged. By default, the period of user activity after which an inactive session is automatically closed is set to 120 minutes. You can change this time period by modifying the value of SessionInactiveLifetimeInMinutes in the web.config file on the Application Server. We do not recommend purging inactive sessions too often. The recommended minimum value is 10 minutes.

If you are using Mozilla Firefox, we recommend that you close tabs rather than the browser window. This issue does not occur with Google Chrome or Microsoft Edge.

Update 2

Technical Information

Release	Part #	Build #	Installer build #	OCRT build #	Release date
Release 3 Update 2	1299/42	12.0.3.4038	171689	16.1.1014.14	2020.07.23

New Functionality

1. Purchase order and receipt processing

This update includes two new out-of-the-box solutions — Purchase Orders and Receipts.

The new Purchase Orders and Receipts solutions include Document Definitions specifically designed for processing purchase orders and receipts, respectively. To create a project using one of these solutions, click **File** > **New Project** and select the desired project type in the **Project Type** field.

To use these solutions, you will need a license that allows you to use the Document Definition associated with that solution.

1.1. Purchase order processing

A purchase order is a document used for ordering goods.

ABBYY FlexiCapture now includes an out-of-the-box solution for processing purchase orders named “Purchase Orders.” To start using the Purchase Orders solution, click **File**>**New Project**, and select **Purchase Order** in the **Project Type** field.

Purchase orders are in many respects similar to invoices, and so are processed using similar algorithms. This means that you will be able to use a database of companies, create training batches for documents originating from a specific company, use country settings for country-specific keywords and tax rates, etc.

To capture purchase order fields, the program uses both FlexiLayouts and machine learning based on neural networks. Both methods are intelligently combined to achieve the best possible results. Any fields that the program fails to capture using its machine learning algorithms will be captured by means of a suitable FlexiLayout.

1.1.1. List of fields

Field	Description
Order number	The number of the purchase order.

Order date	The date when the purchase order was created.
Buyer (group of fields) • Name • BuyerID • VATID • IBAN • Bank code • Bank account • Address • Street • Postal code • State • City • Country	The entity or individual that ordered the goods (i.e. the creator of the purchase order). This is similar to “vendor” on invoices. You can use a database of buyers. To do this, associate your database of buyers with the data set named “Variant.” When no database is used, only the following fields will be captured: • Name • Address
BU (group of fields) • Name • BUID (for database lookups) • VATID • IBAN • Bank code • Bank account • Address • Street • Postal code • City • State • Country	The company that sells the goods (i.e. the recipient of the purchase order). This is similar to “BU” on invoices. You can use a database of business units. To do this, associate your database of business units with the data set named “BusinessUnits.” When no database is used, only the following fields will be captured: • Name • Address
Ship to (group of fields) • Name • Address	The delivery address.
Bill to (group of fields) • Name • Address	The address to which to send an invoice.
LineItems	PO line items.
Position	The number of a line item
Article number BU	The article number in the business unit’s database. You can use a database of article numbers. To do this, associate your database of article numbers with the data set named “Articles“.

Article number	The article number in the buyer's database.
Description	A description of the goods.
Quantity	The quantity of the ordered goods.
Unit of measurement	The unit used for counting, weighing, etc. the goods.
Unit price denominator	The number or amount of goods in each unit. (If the goods are priced per package or per crate, for example, the *unit price denominator* indicates the weight of the package, the number of bottles in the crate, etc.)
Unit price	The price charged for a unit of goods.
Total	The total amount payable for the line item.
Delivery date	The date when the goods should be delivered.
Total	The total amount payable for the entire purchase order. Unlike invoices, this field is not obligatory on purchase orders.
Currency	The currency to be used for payment.

Tax Details (group of fields)
• TotalTaxAmount
• TotalNetAmount
• Tax Group
  ○ Net amount
  ○ Tax amount
  ○ Tax rate
	This group of fields is used for looking for tax amounts on purchase orders and is disabled by default.
	If you want the program to look for tax amounts on purchase orders, open the properties of the Document Definition named "Orders," click the **Document Definition Settings** tab**,** then click **Additional Fields and Features** > **Edit**..., and select the **Taxes** option.

Delivery date	The date when the goods should be delivered.
Additional costs	Any additional charges (e.g. the cost of delivery).

1.1.2. Supported countries and languages

The neural networks used by the machine learning feature were trained on marked up purchase orders from the following countries and in the following languages:

- USA (English)
- Germany (German)
- France (French)
- Spain (Spanish)

For the countries and languages listed above, a number of settings and keywords have been added to enable data capture from purchase orders from these countries. There are also special localization files that enable the display of field and rule names in the appropriate language. The required localization files are installed when you choose the desired GUI language in the installation wizard.

1.1.3. Validation rules

Below are listed the validation rules available for the "Orders" Document Definition. You can enable or disable any of these rules. Many of these rules are similar to those used for validating invoice data.

1. Rules for buyer lookup
   a. Buyer must be present in database
   b. Buyer details must be added to database
2. Rules for business unit lookup
   a. Business unit must be present in database
   b. Business unit details must be added to database
3. The "Total: Amount Check" rule adds up the line item prices and additional costs and compares the result against the number in the "Total" field.
4. The "Separate currency from amount in Total field" rule separates the currency sign from the number if OCR fails to differentiate the currency name from the digits.
5. Currency check
   a. Set currency to CAD or USD – selects the appropriate currency if Canada or USA is specified as the country of the BU.
   b. Detect national currency – detects the currency based on the country specified for the BU.
   c. Copy currency form LineItems – copies the currency from the line items (if detected).
   d. Check currency value – checks the currency if the "Total" field is not empty.
6. Line item check
   a. Check item amount – checks the amount in each line item (quantity multiplied by price minus discount must be equal to the amount stated on the purchase order)
   b. The "Separate currency from amount in Total field" rule separates the currency sign from the number if OCR fails to differentiate the currency name from the digits: "Separate currency…" for the columns with amounts.
   c. Quantity is required – makes sure that the quantity of goods has been detected.
   d. Hide currency if total is empty – removes the currency if no "Total" field has been detected.
   e. Check line item currency value – checks the currency value.
7. Rules for article number lookup
   a. With BUID check: article number must be present in database
   b. Without BUID check BUID: article number and BU must be present in database
   By default, the second rule is enabled. If you do not need a BUID check and want to look up article numbers without BUID filtering, disable the second rule and enable the first rule instead.
8. The "Delivery date must be later than order date" rule checks if the detected "Delivery date" is later than the detected "Order date."
9. Tax rules have been transplanted from the invoice solution but are disabled by default. These rules will be enabled automatically if you enable the "Taxes" option. To enable the "Taxes" option, open the properties of the Document Definition named "Orders," click the **Document Definition Settings** tab**,** then click **Additional Fields and Features** > **Edit**... and select the **Taxes** option.

1.1.4. Data sets

The "Orders" Document Definition uses the following data sets:

- Business units – for BU lookups.
- Variants – for buyer lookups. Training batches will be created for each individual buyer and will be associated with the appropriate VariantID.
- Articles – for article number lookups.

To associate a database with a corresponding data set, open the properties of the "Order" Document Definition, click the **Data Sets** tab, select the desired data set, and click **Setup**. In the dialog box that opens, specify a connection string

to be used by the program for connecting to the database, and configure database columns and lookup options, refresh times, and some other options.

For details, please refer to Connecting vendor and business unit databases in the User's Guide (this article provides instructions for invoice capture, but they are equally applicable to purchase orders).

1.1.5. Quality tests

The table below shows the quality of extraction for header fields and line items on purchase orders from the supported countries. For this benchmark, we used high-quality images of purchase orders from the main supported countries. The quality you obtain on low-resolution scans and photos may be lower. The field is deemed correctly detected when its region is correct (i.e. when its region exactly matches the reference region).

Country	Order date	Order number	Total	Line items	
				Rows	Qty, Unit price
DE	88.4%	94.6%	95.8%	85.1%	83.7%
ES	96.4%	66.7%	92.8%	65.2%	65.2%
FR	89.5%	67.2%	87.1%	83.1%	78.6%
US	80.9%	96.1%	87.6%	65.7%	63.2%

1.2. Receipt processing

A receipt is a document that confirms a payment for goods or services.

ABBYY FlexiCapture now includes an out-of-the-box solution for processing receipts named "Receipts." To start using the Receipts solution, click **File** > **New Project**, and select Receipt in the **Project Type** field.

Data from receipts is captured using machine learning alone, no FlexiLayouts are used.

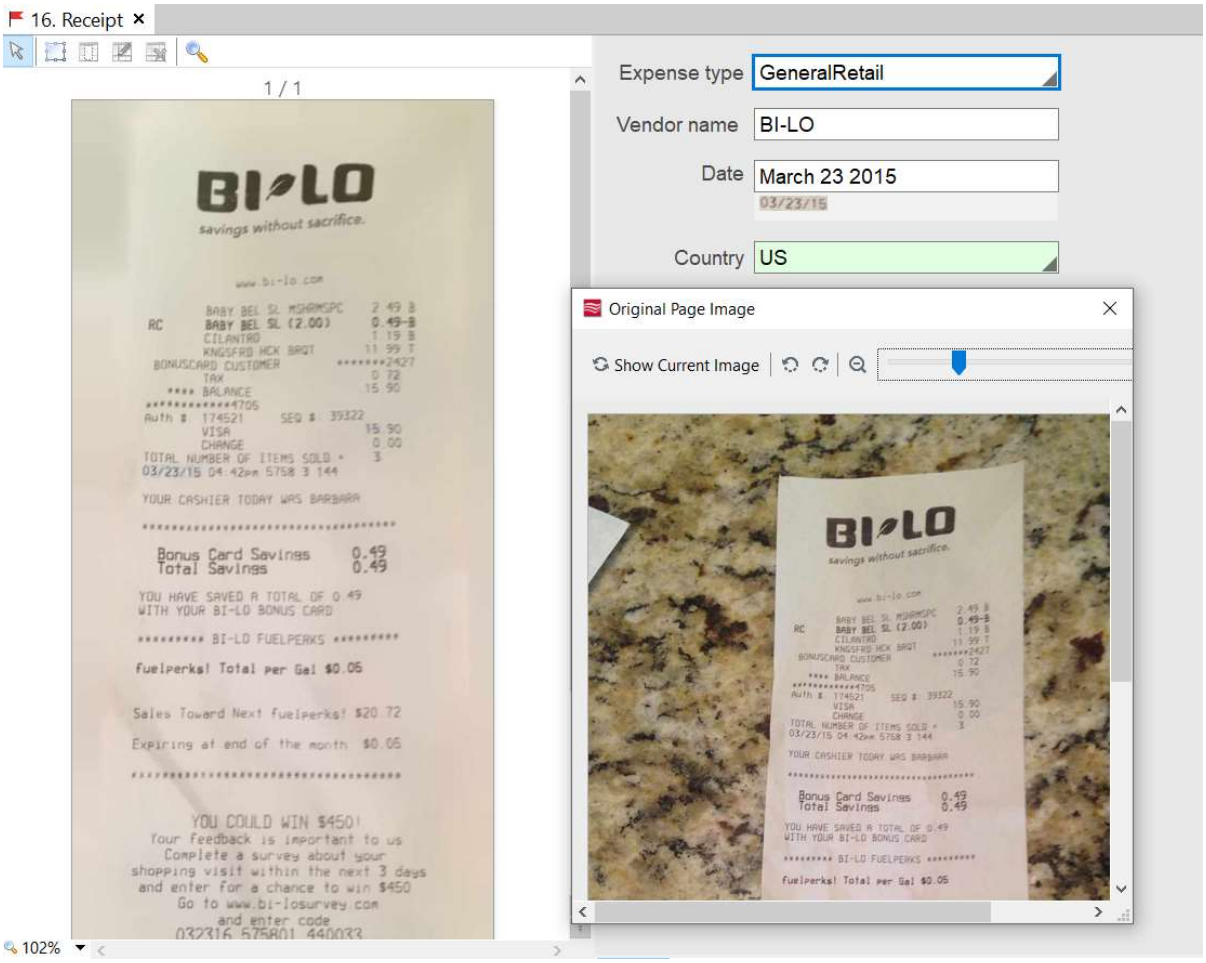
1.2.1. Built-in image processing

Images of receipts are often available as photos taken with phone cameras and require some preliminary processing before their data can be reliable captured. The program will first classify the receipt images into scans and photos, because each type of images requires different corrections. Autocrop and line straightening are only performed on photos, where they greatly improve the quality of field detection.

If you want to be able to see the original images, enable the **Store original image during processing** option when loading your images. This option can be enabled in the project properties or in the batch type properties (on the **Image processing** tab), when creating an image import profile, or when loading images manually.

If the original image is available for a processed image, you will be able to review it during verification (right-click the processed image and then select **Image** > **Show original image** on the shortcut menu).

An example of a processed image and its original:



1.2.2. List of fields

Field	Description
Expense type	The type of expense is detected by a pretrained classifier. The classifier was trained on the following types of documents: 1. Airfare – documents confirming flight payments 2. CarParking – documents confirming payment of car park charges 3. CarRental – documents confirming car rental payments 4. GasolineStation – documents confirming motor fuel payments 5. GeneralRetail – documents confirming payments at supermarkets 6. Hotel – documents confirming hotel payments 7. Restaurant – documents confirming restaurant payments 8. Taxi – document confirming taxi ride payments 9. TollRoad – documents confirming road toll payments 10. Transport – documents confirming fare payments
Vendor name	The name of the organization that accepted the payment and issued the receipt.
Date	The date when the receipt was issued.
Country	The country where the goods or service was purchased and paid for. When verifying a recognized receipt, you can correct the country if necessary by selecting the appropriate country from a drop-down list.
City	The city where the goods or service was purchased and paid for.

Currency	The currency of the payment. When verifying a recognized receipt, you can correct the currency if necessary by selecting the appropriate currency from a drop-down list.
Total	The total amount of the receipt.
Tax	The total amount of tax payable on the purchase.
Taxes • Tax rate • Tax amount	Tax broken down into different categories if several taxes have been detected. This will be displayed as multiple lines of a repeating group. The total tax amount will be recorded in the "Tax" field.
The following fields will only appear on the data form for Hotel receipts:	
Check-in date	The day when the customer checked in.
Check-out date	The day when the customer checked out.
The following fields will *not* appear on the data form by default. To display these fields, you have to edit the Document Definition named "Receipts," enable the **Show on verification** option, and specify the location of each field on the data form.	
Time	The time of the payment.
Address	The vendor's address.
Phone	The vendor's phone number.
Subtotal	Amount before tax.
VatID	The taxpayer number of the vendor's company.
TaxName	The name of the tax.
TaxCode	The code of the tax.
LineItems • SKU • Description • Quantity • Price • Amount	Line items with the goods or services purchased.

1.2.3. Supported countries

The table below lists the countries whose receipts can be captured by means of ABBYY FlexiCapture. The country classifier was trained on receipts from all of these countries. The neural network used for data capture was specifically trained for some of these countries (marked with a "Yes" in the "Network trained column"). For the other countries, a generic algorithm will be used.

Country	Code	Network trained	Main currency

USA	US	Yes	USD
Canada	CA	Yes	CAD
Germany	DE	Yes	EUR
Italy	IT	Yes	EUR
Spain	ES	Yes	EUR
France	FR	Yes	EUR
Russia	RU	No	RUB
United Kingdom	GB	Yes	GPB
Australia	AU	Yes	AUD
Singapore	SG	No	SGD
Netherlands	NL	No	EUR
Turkey	TR	No	TRY
Poland	PL	Yes	PLN
Sweden	SE	Yes	SEK
Finland	FI	Yes	EUR
Austria	AT	No	EUR
Switzerland	CH	No	CHF

If you intend to process receipts from only one particular country, you can specify the two-letter code of that country (as indicated in the "Code" column above) in the **fc_Predefined:ReceiptPredefinedCountries** registration parameter. This will turn off the country classifier, and the program will use instead the country you have specified in the registration parameter. You can specify multiple country codes, separating them with ";", in which case the country classifier will give priority to the countries you have specified in the registration parameter.

1.2.4. Validation rules

Below we list the validation rules available for the "Receipts" Document Definition. You can enable or disable any of these rules.

1. Calculate tax amount – calculates the total tax amount.
2. Expense type from list – allows you to select an expense type from a drop-down list.
3. Choose country form list – allows you to select a country from a drop-down list.
4. Choose currency form list – allows you to select a currency from a drop-down list.
5. Show only for hotel – two rules that display the check-out date and check-in date only for hotel receipts.
6. Check if amount equals total – adds up the amounts in the line items and compares the sum against the total amount of the receipt. This rule and the line items table are disabled by default. When enabling the display of line items, you may want to enable this validation rule as well.

1.2.5. Quality tests

The table below shows the quality of extraction for header fields on receipts from the supported countries. For this benchmark, we used receipts, both photos and scans, from the main supported countries.

Country	Percentage with 3 correct fields	Date	Total	Currency	Country detection
Australia	76.7%	84.5%	88.27%	98.22%	98.16%
Canada	82.6%	91.11%	88.93%	97.27%	97.77%
Germany	84.1%	90.61%	93.89%	95.05%	94.08%
Finland	84.2%	88.97%	94.18%	98.37%	98.26%
France	86.3%	91.66%	93.41%	99.23%	98.82%
Italy	84.2%	92.79%	89.38%	99.4%	99.2%
Poland	79.5%	85.75%	91.72%	98.39%	97.93%
Spain	77.9%	89.02%	84.53%	96.84%	95.34%
Sweden	80.1%	92.04%	86.22%	99.85%	99.85%
UK	78.4%	87.04%	88.21%	100%	99.27%
USA	87.1%	92.01%	93.44%	99.09%	99.09%
Non-trained control countries					
Netherlands	88.3%	92.48%	94.44%	99.59%	99.51%
Singapore	77.1%	87.79%	88.80%	94.59%	95.42%
Turkey	78.9%	94.50%	91.74%	91.74%	94.5%

1.2.6. Using a data set to detect the vendor

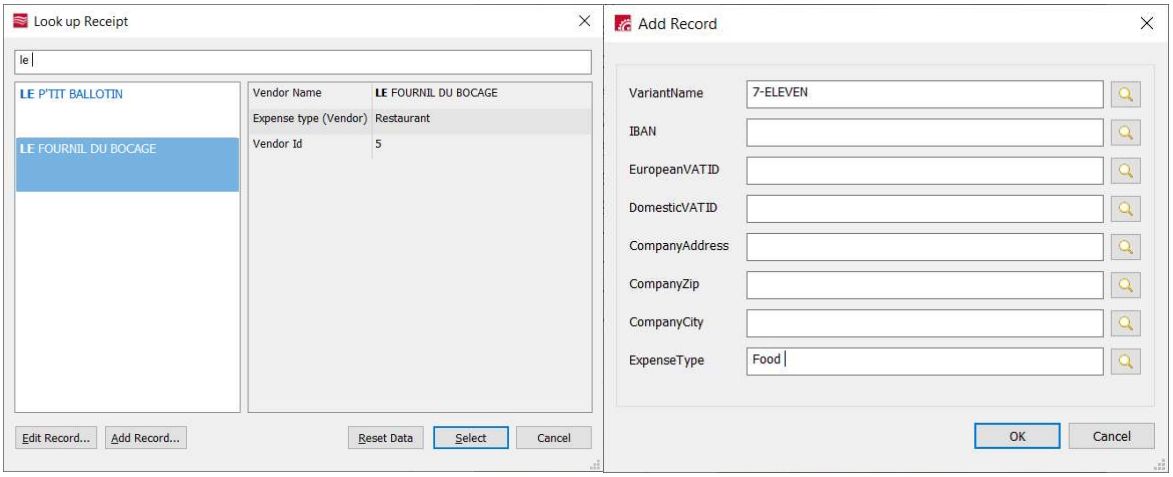
Sometimes, vendors can be more reliably detected if you accumulate and use a data set of possible vendors. Using this data set, you will also be able to specify custom expense types for certain vendors. For example, if company employees usually buy office supplies and stationery from a particular vendor, you can specify “Office Supplies” as the expense type for this vendor. Now for all captured receipts from this vendor, the expense type will be automatically changed to “Office Supplies.”

The fields and rules for data set checks are disabled by default. To enable them, complete the steps below:

1. In the Document Definition Editor, enable these fields: VendorID_Dataset, VendorName_Dataset, VendorExpenseType_Dataset, and FinalExpenseType.
2. Enable these rules: Vendor database check, Copy expense type, Copy company name.
3. To be able to look up vendors during verification, add a button onto the data form (right-click on the form and select **Insert > Button** on the shortcut menu). Then, on the **Format** tab, select the **Database lookup** type and the **Vendor database check** rule.
4. Publish your Document Definition.

Names of vendors will be automatically added to the data set each time the program detects a new name on a receipt. As your data set of vendors grows, the program will be able to detect vendors more reliably.

To edit the vendor data set or to find a particular vendor, click the button that you added onto the data form and then click the **Edit record**, **Add record**, **Reset data** or **Select** button, depending on the action you want to perform.



When adding or editing a record, you can specify an expense type typical of the given vendor. Next time you capture a receipt from this vendor, this expense type will appear in the **Expense type (Final)** field.

1.2.7. Limitations

1. The program cannot capture data from multiple receipts scanned together as one page. To process this type of scan, you need split the page into separate receipts first. On Scanning and Verification Stations running on Windows machines, this can be done using the **Split** tool.
2. Hand-written amounts or receipts executed entirely by hand cannot be captured.

1.3. New features in ABBYY FlexiCapture for Invoices

1.3.1. ABBYY FlexiCapture for Invoices can now operate without a database

Machine learning will now be used to detect company names and addresses if no database of companies is used.

Machine learning will be used to detect the following companies and addresses:

- Vendor and BU on invoices
- Buyer, BU, ShipTo, and BillTo on orders (see the “Purchase orders” section for details).

Quality of BU and vendor detection (measured on invoices):

	BU name	BU address	Vendor name	Vendor address
USA	81.52%	79.90%	81.25%	85.91%
Germany	99.20%	89.85%	85.89%	90.09%

1.3.2. Selecting country settings working without database

To select the right country settings, the program needs to know the vendor or BU country. The program will attempt to pick out the right country from the database of companies. If neither the vendor nor the BU is found in the database, or in case of working without database, the program will use the country designated as “the default country.” During tax search the system is able to detect a country automatically.

1.3.2.1. Country detector for tax search

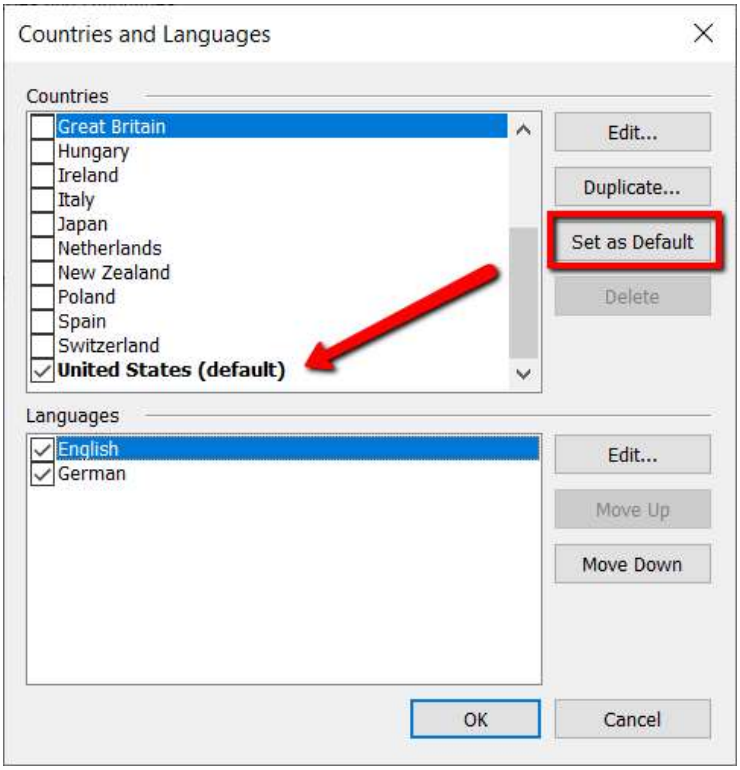
Knowing the country of a vendor or BU enables the program to select the right tax rates and keywords for subsequent detection of tax amounts. If a database of companies is used, the program will take the vendor and BU countries from the database. If no database is used, the countries will be detected automatically.

1.3.2.2. Specifying a default country

The program will use the languages specified for the default country when looking for keywords pointing to tax amounts. The automatic country detector described in 1.3.2.1 above is launched later and will only affect tax detection, while the keywords will be selected based on the country designated as the “default country.”

To specify a default country, open the Document Definition properties, click the **Document Definition Settings** tab, click **Edit…**, and specify a country in the **Countries and languages** section.

There must always be a default country. If you clear a check box next to the current default country, the program will automatically choose the next country in the list as the default country.



Alternatively, you can specify a default country in the **DefaultCountry** parameter in the **document_settings.xml** file. In this case, the same country will automatically be selected in the GUI.

A default country can be specified for invoices and purchase orders.

In the newly added projects, the following are specified as the default countries:

- InvoicesEU – Germany
- InvoicesUS – USA
- InvoicesCA – Canada
- InvoicesJP – Japan
- InvoicesANZ – Australia
- InvoicesES – Spain
- Orders project – USA

1.3.3. New country projects

Two new country projects have been added for Japanese and Spanish invoices.

1.3.3.1. Japanese invoice project

The Japanese invoice project is named “InvoicesJP” and can be selected when creating a new project. This project uses a Document Definition designed specifically for handling Japanese invoices.

Country	Invoice number	Invoice date	Total	Line items	
				Rows	Qty, Unit price, Total
Japan	77.44%	93.20%	71.26%	73.76%	62.48%

1.3.3.2. Spanish invoice project

The Spanish invoice project is named “InvoicesES” and can be selected when creating a new project. This project uses a Document Definition designed specifically for handling Spanish invoices. This dedicated Spanish project has been created for cases when multiple taxes (e.g. IVA and IRTF) are calculated using the same taxable base. For cases like this, the data form in this project includes some extra fields that are not available in the European invoice project.

The Spanish invoice project is intended for customers that process mainly native Spanish invoices. For international companies that receive invoices from all over Europe and only rarely encounter Spanish invoices of the type described above, the European invoice project should be preferred.

1.3.4. New country settings

Country settings and keywords have been added for the following countries:

- Lithuania
- Estonia
- Latvia
- Slovakia
- Norway
- Finland
- Sweden
- Denmark
- Portugal
- Turkey

All of these countries are disabled by default. If you intend to process invoices from any of these countries, enable them in any of your regional projects: open the Document Definition properties, click the **Document Definition Settings** tab, then click **Edit…** and in the **Countries and languages** section, select the desired countries.

2. ABBYY FlexiCapture for Enterprises

2.1. Online protection

When license data is stored locally and is tied to the hardware of the machine on which the license was activated, the following issues may be encountered in cloud environments:

- For a failover setup, two or more licenses are required.
- If a virtual machine malfunctions, the administrator has to restore the licenses manually.
- When upgrading hardware behind a virtual machine or moving a virtual machine to another host, the licenses have to be reactivated.

To address the above issues, a new type of license has been introduced. For this new type of license, termed “online license,” all of its data (including expiration date, page counter, allowed features, etc.) is stored on ABBYY’s cloud

servers. The License Server only stores a temporary copy of this data, which is periodically synchronized with the data stored on ABBYY's cloud servers.

The following synchronization parameters can be modified:

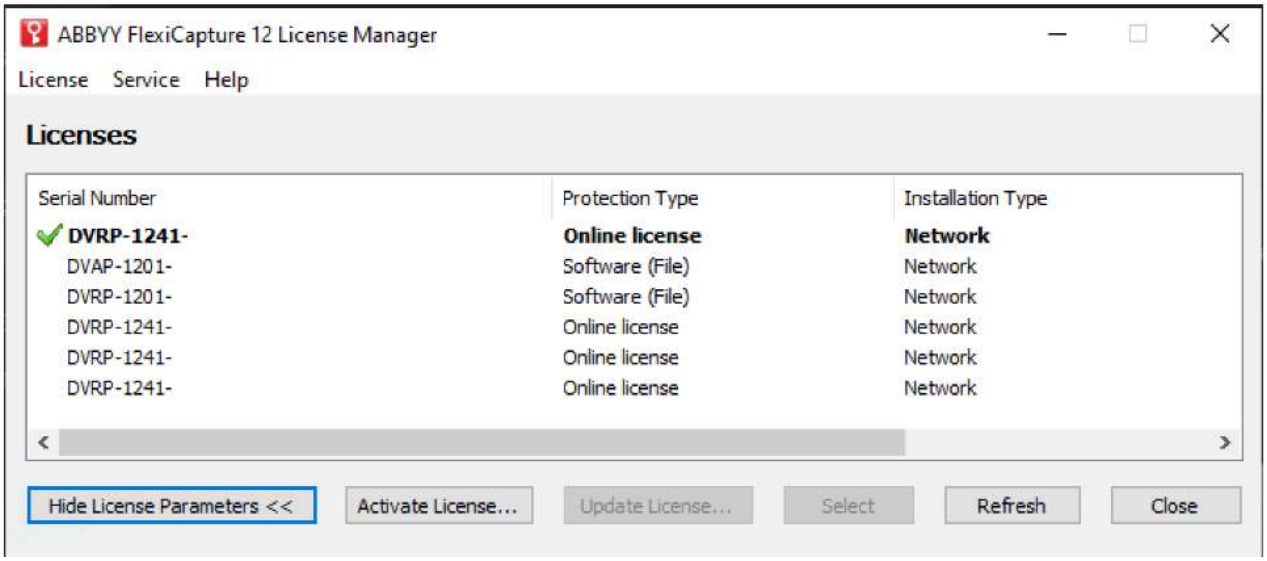
- How often to synchronize the license data
- How long to keep on working after connection to the ABBYY cloud server has been lost

In order to synchronize the license data, the License Server must be connected to the Internet. All traffic between the License Server and ABBYY's cloud servers is encrypted.

No changes have been made to the activation process. Administrators will still need to click the **Activate License** button in the License Manager, enter the serial number, and follow the instructions of the activation wizard.

Important! The License Server must be connected to the Internet both when activating an online license and when working with the program.

Once activated, the license will appear in the list of licenses and will be identified as "Online license" in the "Protection Type" column.



An online license can be activated only once.

An online license can be accessed from different machines. This may be necessary when deploying a cluster or multiple instances of ABBYY FlexiCapture 12, when migrating to a new environment, or when recovering from a system crash.

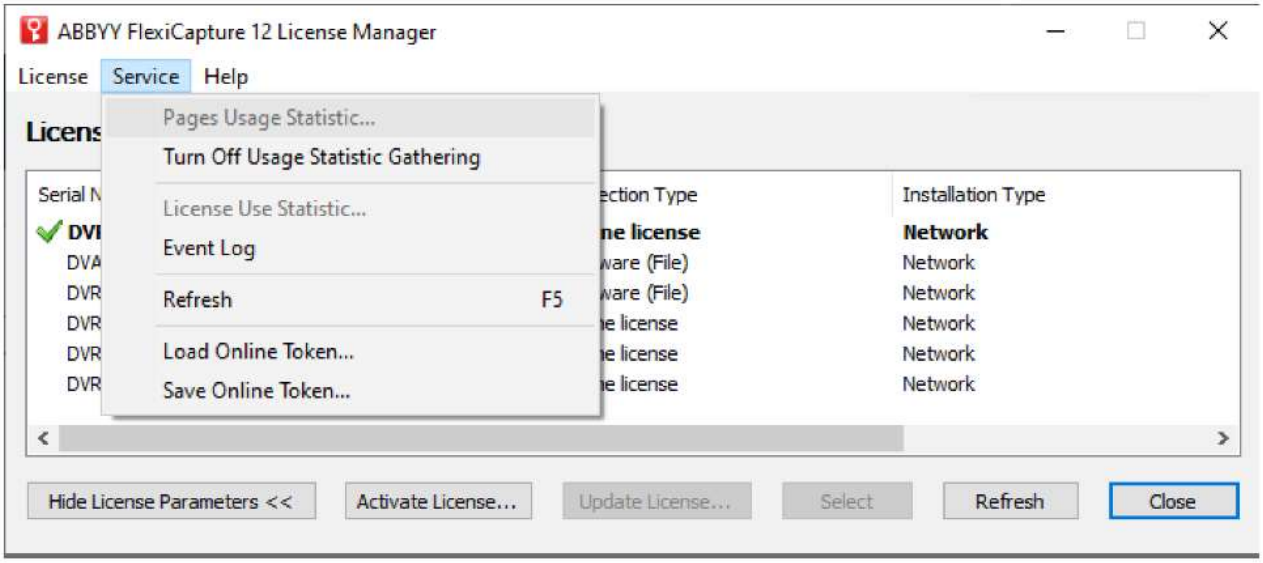
To access an online license on the ABBYY cloud servers, the administrator must use a license token generated for that license. A license token is generated when activating a license and is stored locally in an encrypted format.

To access an online license from another machine:

1. On the machine for which the license was originally activated, save the license token to a file.
2. On the machine from which you want to access the license, load the license token obtained in step 1.

Note: An online license can be accessed from any number of machines.

The corresponding license will appear in the list of available licenses on the new machine. Now you can select this license and make it active.



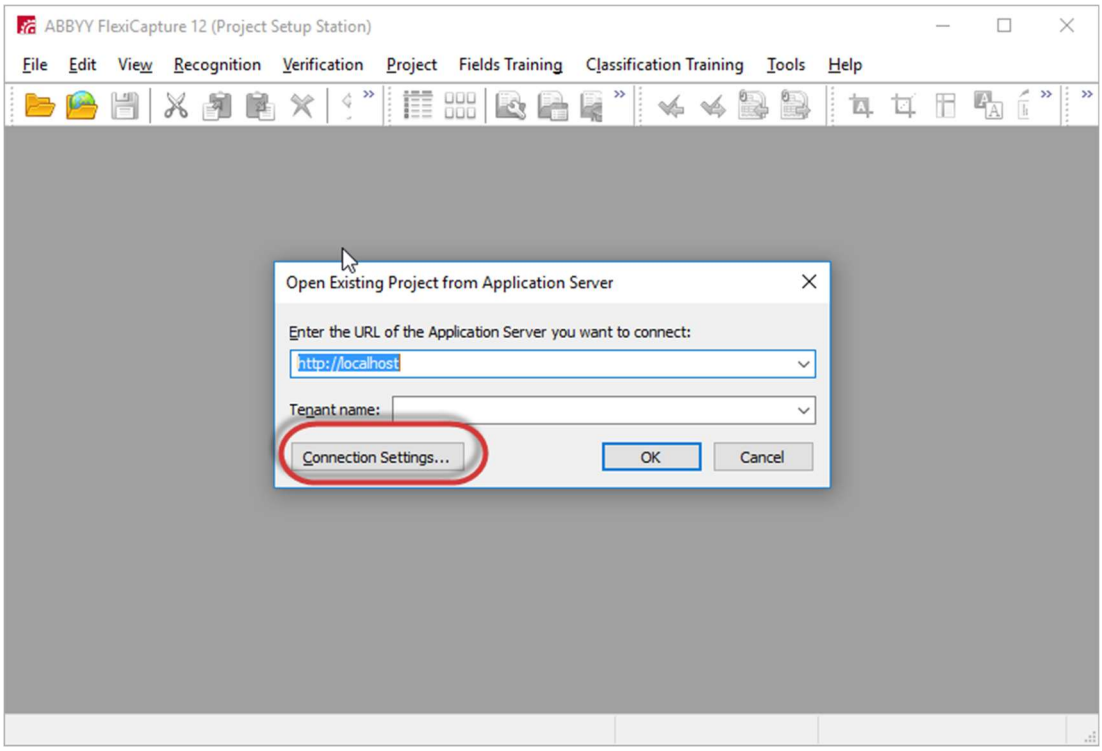
You can use LicenseManager.Console.exe to automate the following operations: activating a license, saving a license token to a file, loading a license token, and making a license active. For details, see [Managing your licenses using the command-line console](#).

2.2. Proxy support for desktop stations

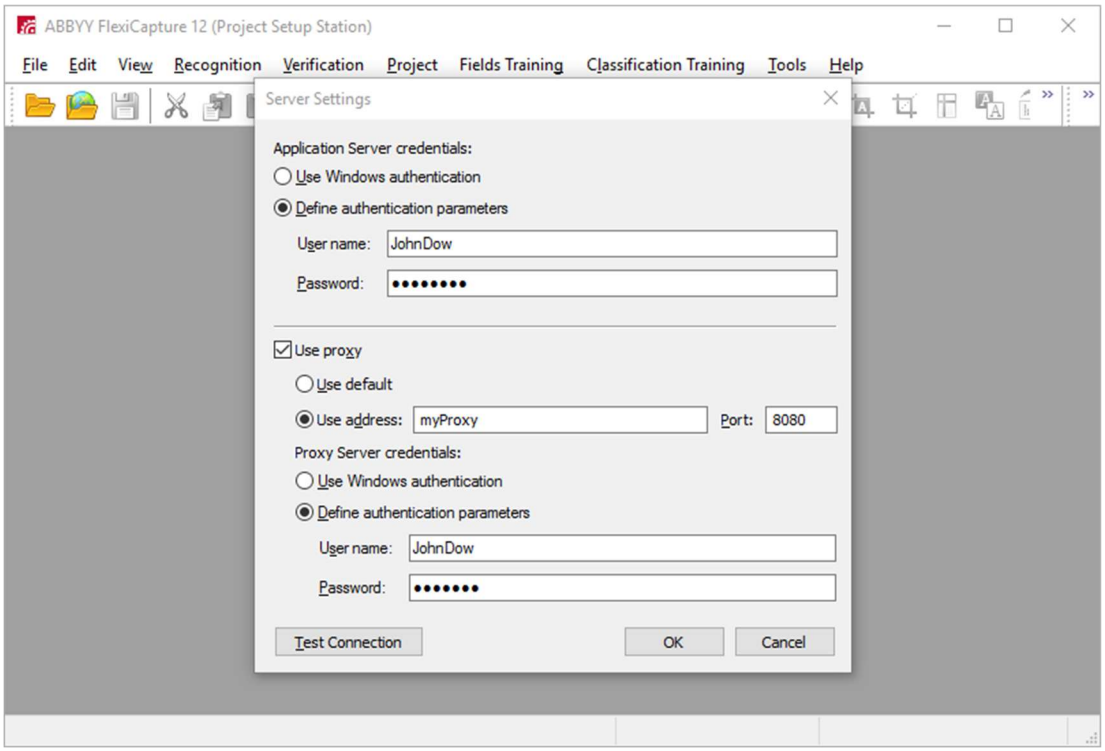
In some installations, the remote Application Server may be deployed behind a proxy (e.g. for load balancing), so in some cases additional authentication may be required in order to get through to the Application Server.

Now you can specify the necessary proxy and authentication parameters on a desktop station. To do this:

1. In the dialog box where you enter the URL of the Application Server, click the **Connection Settings...** button.



2. The **Server Settings** dialog box will open.



3. Select the **Use Proxy** option.
4. Specify the required proxy and authentication parameters.

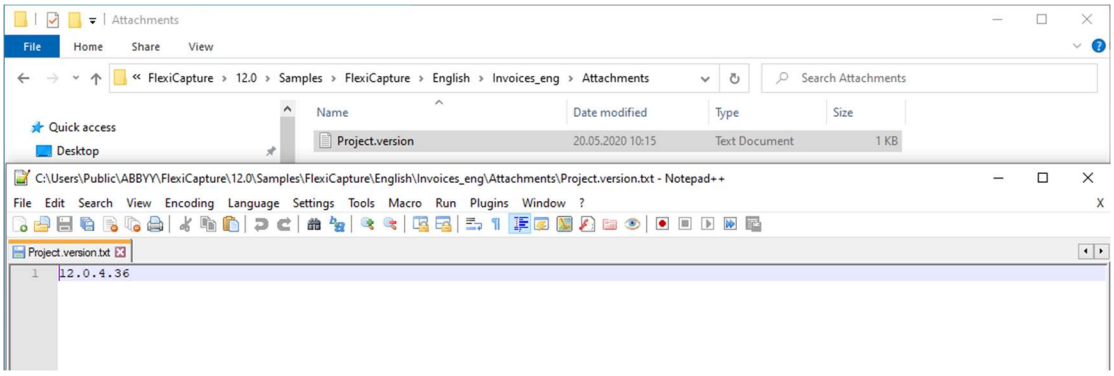
2.3. Microsoft SQL Server 2019 support

ABBYY FlexiCapture 12 Release 3 Update 2 can use Microsoft SQL Server 2019 to store processing data.

2.4. Project version is now saved to file

Previously, users could not easily find out the version of ABBYY FlexiCapture that had been used to create a particular project. The only solution was to try and open the project in different versions of ABBYY FlexiCapture and then update the project to the latest version.

Starting from this Update 2, a separate file is created that contains the version of your project. This file is named “Project.version” and can be found in the Attachments folder. This file is created when you create a new project, when you update an old project, or when you download a project from the server.



2.5. Tasks from unupdated projects no longer sent to Exceptions

When updating their installed copies of ABBYY FlexiCapture, users need to stop the processing of all documents, update ABBYY FlexiCapture, and then update all the projects on the server (for details, please refer to Updating ABBYY FlexiCapture 12 in the Administrator’s Guide). Updating a large number of projects may take a considerable amount of time.

Previously, when users resumed the processing of tasks before all of the projects had been updated, any automatic tasks from the unupdated projects ended up at the Exceptions stage and had to be returned into processing manually.

Starting from this Update 2, if you resume the processing of tasks with some of the projects still waiting to be updated, tasks from these unupdated projects will not be sent to Exceptions but will remain in their respective queues until the update completes.

You can also use FCAdminTools to find out which projects are still waiting to be updated. For details, please refer to CheckProjectsVersion command in the Administrator's Guide

2.6. Using built-in authentication for the Processing Server and Processing Stations

The Processing Server and Processing Stations use Windows authentication to access the Application Server. Authentication is carried out using the account under which the service is running.

In cloud environments without domains, there is no centralized storage location for account credentials, and one machine will not trust a local account from another machine, preventing that machine from authenticating with that local account.

Starting from Update 2, you can force the Processing Server and Processing Stations to use ABBYY FlexiCapture's built-in authentication (see the instructions below).

For the Processing Server:

1. On the Application Server, create an account and grant it Processing Server permissions (see Managing users).
2. In the console, run this command:
   set BasicAuthString=<username>:<password> (see Description of Processing Server commands)

For a Processing Station:

1. On the Application Server, create an account without any permissions.
   Note: Required permissions will be granted to this account automatically by the Processing Server.
2. In the console, run this command:
   set BasicAuthString=<username>:<password>.

Before specifying the authentication parameters, you must open the Processing Server Monitor and stop the Processing Server and the Processing Station. If incorrect authentication parameters are specified, the Processing Server or the Processing Station won't start and an error message will be displayed saying that the Application Server is inaccessible.

You can use the **view** command to check which type of authentication is used. If built-in authentication is enabled, the **view** command will return BasicAuthString=username:password

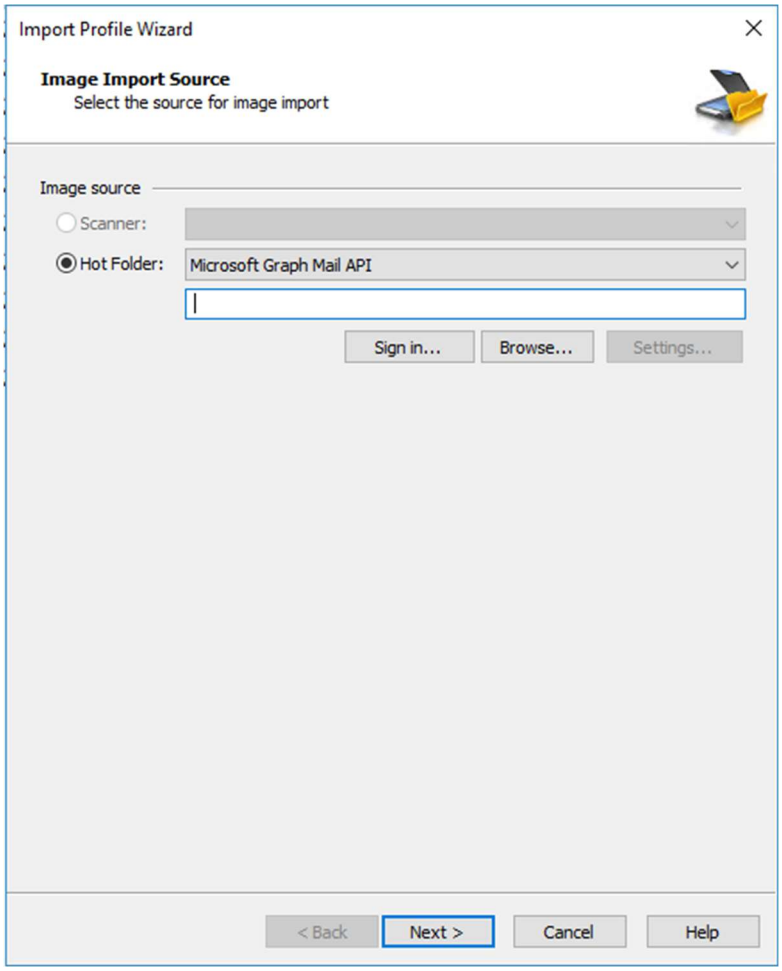
The password is stored and displayed in an encrypted format.

To revert to Windows authentication, run this command: set BasicAuthString=

2.7. Import from Microsoft 365 via Microsoft Graph with OAuth 2.0 authentication

A new import option is available starting from this release – import rom Microsoft 365 via Microsoft Graph. Authentication is performed using OAuth 2.0, which enables applications to access each other's data without revealing their users' credentials. Instead, username and password tokens are used, providing a more secure connection compared to basic authentication.

A technical preview is available for mail import from Microsoft Graph (the recommended method of working with Exchange 365 is via the REST API). To access the Microsoft Graph import settings, modify the following registry key on the machine where the Project Setup Station is installed:
HKEY_CURRENT_USER\Software\ABBYY\FlexiCapture\12.0\Options\EnableBetaGraphMailImport = true.



This will unlock a new image import source, Microsoft Graph Mail API.**Known issues and limitations**

http://rm/view/907620 - OAuth wizard: The Project Setup Station becomes unresponsive when quitting the OAuth wizard while editing a Microsoft Graph import profile (and sometimes when launching the OAuth wizard).

3. Using Natural Language Processing (NLP) for data extraction

3.1. Generalized NLP model (cross-vendor training)

Previously, the only available method of training NLP models was as follows:

1. Set up NLP models in your Document Definition.
2. For field extraction training, create a training batch. For this training batch, specify the Document Definition, section name, and variant or vendor name.
3. Enable the NLP option for the training batch.

If your Document Definition included multiple variants or vendors, you had to create a training batch for each.

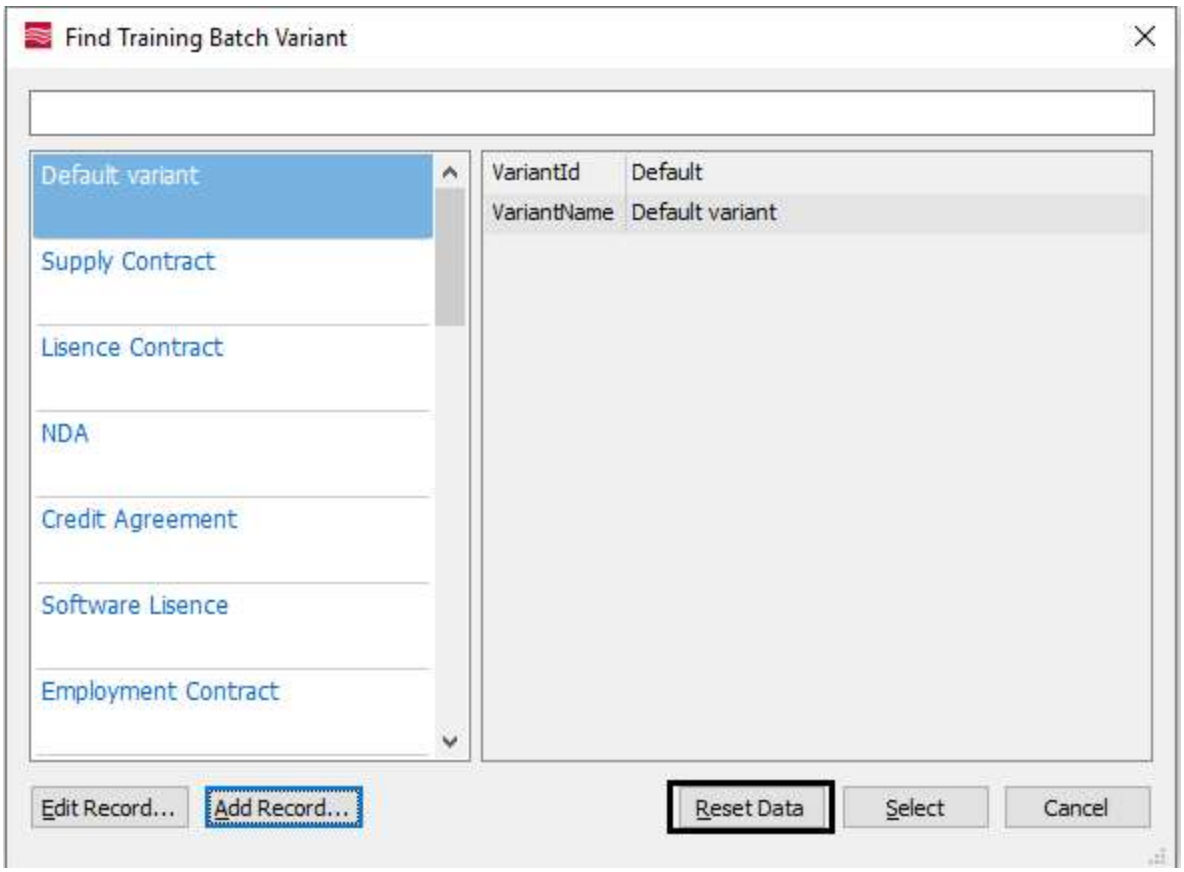
As each variant or vendor has similar-looking documents, the model can be trained to a very high standard, provided there is a sufficient number of documents in the training batch.

This method, however, has two limitations:

1. If your Document Definition includes a large number of variants or vendors, you need to create a lot of training batches.
2. You may not have enough documents to train some variants or vendors.

At the same time, NLP models are only good at generalizing contexts if your training batch is representative enough.

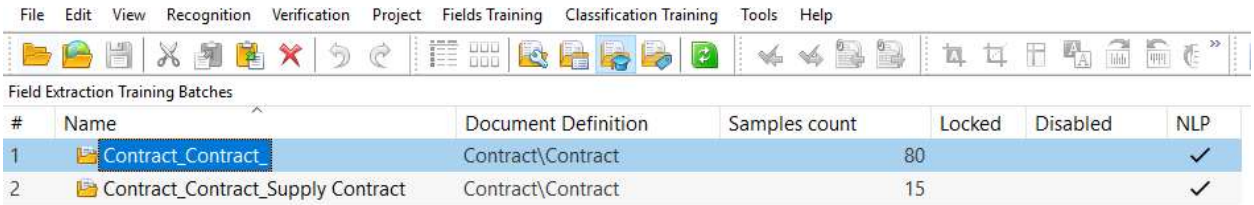
To overcome the above limitations, now you can have a training batch that is not associated with any variant or vendor. To associate your training batch with a section, click **Change Variant… > Change Vendor** and then click **Reset Data**.



If a variant or vendor has an NLP training batch associated with it, the program will use the models that were trained on this batch for this variant or vendor. If a variant or vendor has no NLP training batch associated with it, the program will use the models trained on the “generic” training batch associated with the section.

When to use a training batch for a specific variant or vendor?

A separate training batch may be useful if the generic model delivers poor results for a particular variant or vendor. This may be due to an insufficient number of documents for a particular variant or vendor, or due to a conflict between similar contexts. For example, one variant requires that certain contract terms be extracted from the preamble, while another variant has these terms arranged in a tabular format at the start of the document.



3.2. Training based on feedback from verification operators

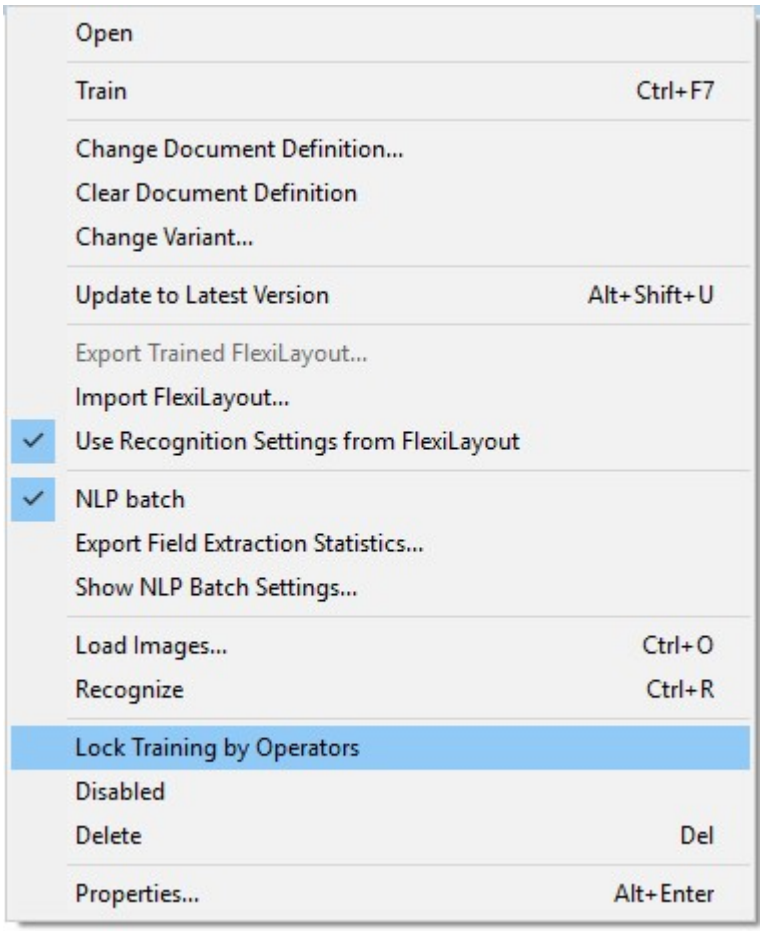
The quality of data extraction can be improved through additional training of NLP models by operators. If the program fails to detect certain fields or mistakes one field for another, the verification operator can indicate the correct field and retrain the NLP model. The program will then use the retrained model for more accurate data extraction.

There are two ways to initiate the training of an NLP model during verification:

- Add a training stage after the verification stage. Training will start when the conditions specified for the training batch are met.
- Manually send documents to the training stage. To do this, right-click the working batch or a document in the working batch and select **Train** on the shortcut menu.

When training is initiated, ABBYY FlexiCapture will automatically create a generic training batch in the list of training batches (if it does not contain one already). This generic training batch will be associated with the appropriate section. Now all documents related to a specific Document Definition will be copied into this batch, regardless of their variant.

Important! Make sure that the training batch is not locked for training by operators.



Some of the documents will be set aside for testing the new model.

Once the training is finished, the program will test the existing model and the new model on the documents it has set aside for testing. If the overall performance of the new model is not worse than that of the existing model, the existing model will be replaced with the new one. Otherwise, the new model will be rejected.

If the list of field training batches contains an active NLP training batch for a specific variant or vendor, documents for this variant or vendor will be copied into this NLP training batch during training initiated by operators. If there is no active NLP training batch for a specific variant or vendor, the program will copy the documents into the generic training batch associated with the section and use the generic NLP model.

You can specify some additional options for a training batch:

- **Maximum documents in each training batch.** If the maximum number of documents is reached, any new documents added into a training batch will start replacing old documents. This option causes the model to adapt to new documents faster.

- **Maximum percentage of replaced documents.** Indicates the percentage of old documents that can be replaced with new ones during one training session. If the operator attempts to use too many new documents for training, only the allowed percentage will be used.
- **Start training if batch contains more than __ new documents or more than __ % of new documents.** Training will start only when the number or percentage of new documents added into a training batch becomes equal to or greater than the specified value. Otherwise, training won't start and an entry will be added into the task log saying that there are not enough new documents to start training. This option prevents the model from being trained too often. Tasks will not be queued if a sufficient number of documents has accumulated to warrant another training session. Instead, the next training session will be initiated only when the next portion of new documents is added into the batch. If the operator attempts to add too many documents in one go, the excess documents will be ignored. If you encounter this situation, we recommend increasing the maximum number of documents allowed in a training batch and/or the maximum percentage of replaced documents.
- **Percentage of documents to be used for training.** Specifies the percentage of documents to be used for training (and, by implication, for testing). For example, if you limit the percentage of "For training" documents to 70%, the remaining 30% will be marked "For testing."

Newly trained NLP models will be tested and accepted only if they perform better or at least no worse than existing models

Once the training is finished, the program will test the existing model and the new model on the documents it has set aside for testing.

You can review the statistics for the old and the new NLP models in the training log or export them to a CSV file. To export the log for a training batch, right-click the batch, click **Export Field Extraction Settings…** on the shortcut menu, and specify where you want to save the CSV file.

The newly trained model is deemed to be no worse than the existing model if its F-measure over all the fields is at least equal to the similarly calculated F-measure of the existing model.

3.3. Using NLP models created in earlier versions of the program

Starting from Update 2, you can install different versions of the NLP module. After you upgrade to a newer version of ABBYY FlexiCapture (e.g. to Update 2), models trained on earlier versions (e.g. Release 3) will perform the same on the same texts.

If you delete an old NLP module, you will need to retrain your models using the new version of the NLP module.

You can find out the version of the NLP module used for training your current models by exporting the training statistics to a CSV file (see above).

3.4. New out-of-the-box models

New NER model

The table below lists the entities that can be recognized by NER in English, Russian, and German texts. The entities with "Yes" typed in bold have been added in this Update 2.

Entity (only span, w/o parsing)	English	Russian	German
Person	Yes	Yes	Yes
Organization	Yes	Yes	Yes
Location (country, town)	Yes	Yes	Yes
Address	Yes	Yes	**Yes**

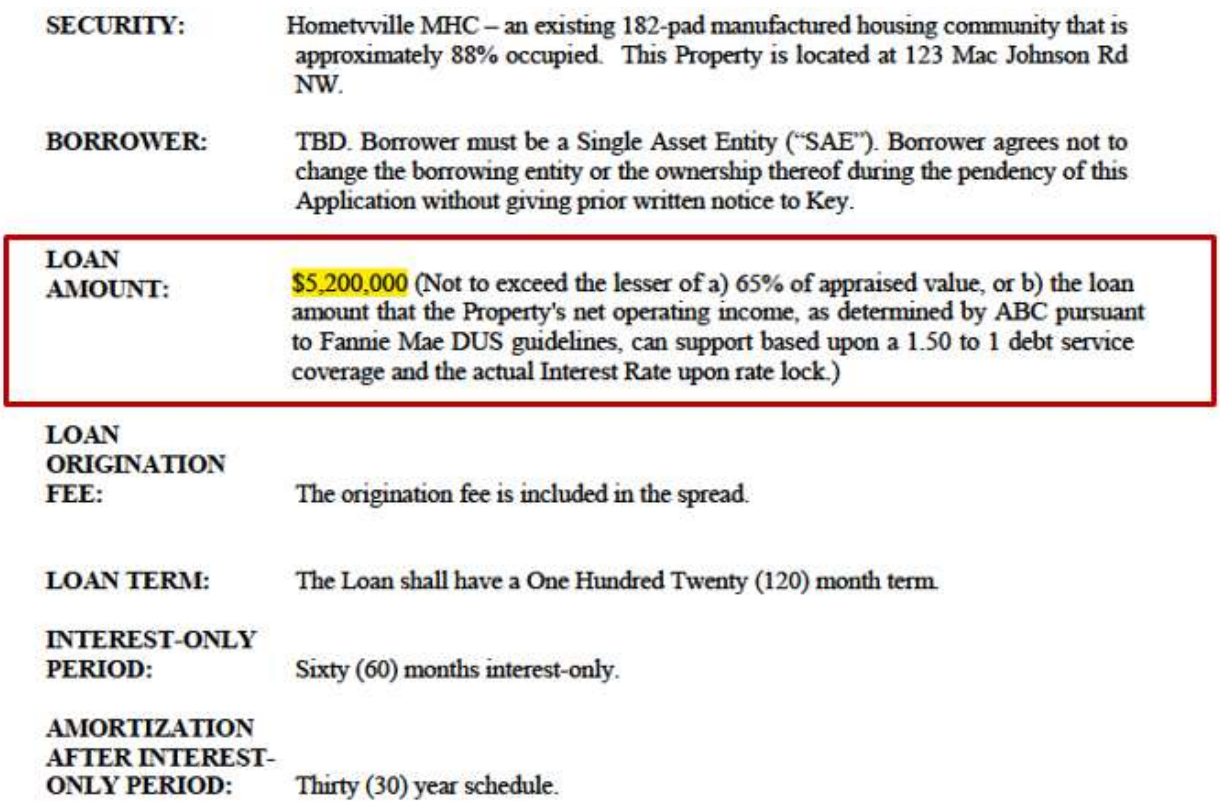
Money	**Yes**	**Yes**	–
Date	**Yes**	**Yes**	–

How to use NER? Quick pilot vs. production use

If you have enough documents and time to work on your project, we recommend using ML, i.e. Segmentation and Extraction. In this case, ABBYY FlexiCapture will select the best method to use for field extraction — you will just need to provide it with a sufficient number of marked up documents.

If you need to create a quick pilot to demonstrate the NLP capabilities, a combination of Segmentation and NER/RegEx/Dictionaries will suffice.

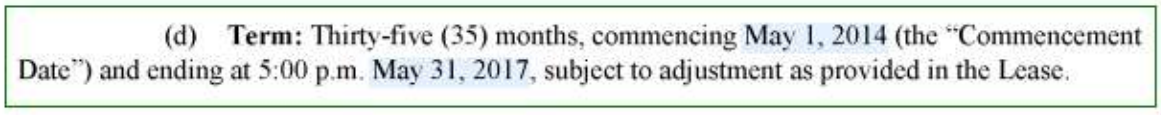
1. Loan Amount



SECURITY:	Hometvville MHC – an existing 182-pad manufactured housing community that is approximately 88% occupied. This Property is located at 123 Mac Johnson Rd NW.
BORROWER:	TBD. Borrower must be a Single Asset Entity ("SAE"). Borrower agrees not to change the borrowing entity or the ownership thereof during the pendency of this Application without giving prior written notice to Key.
LOAN AMOUNT:	$5,200,000 (Not to exceed the lesser of a) 65% of appraised value, or b) the loan amount that the Property's net operating income, as determined by ABC pursuant to Fannie Mae DUS guidelines, can support based upon a 1.50 to 1 debt service coverage and the actual Interest Rate upon rate lock.)
LOAN ORIGINATION FEE:	The origination fee is included in the spread.
LOAN TERM:	The Loan shall have a One Hundred Twenty (120) month term.
INTEREST-ONLY PERIOD:	Sixty (60) months interest-only.
AMORTIZATION AFTER INTEREST-ONLY PERIOD:	Thirty (30) year schedule.

To extract the loan amount from the document shown in the picture above, you first need to configure the segmentation feature to extract the "Loan Amount" segment, and then use the NER feature to extract the amount of money.

2. Lease Commencement and Expiration Dates



(d) **Term:** Thirty-five (35) months, commencing May 1, 2014 (the "Commencement Date") and ending at 5:00 p.m. May 31, 2017, subject to adjustment as provided in the Lease.

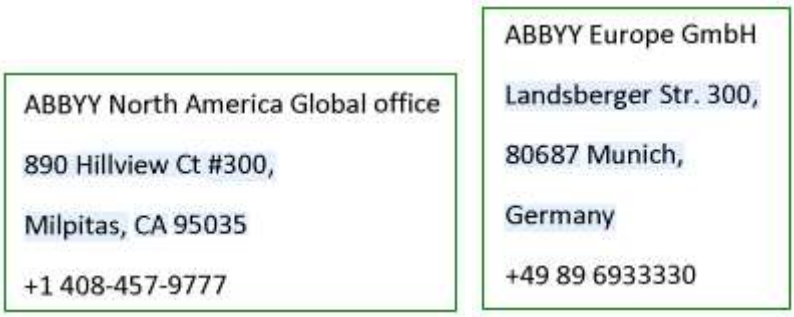
To extract the lease commencement and expiration dates from the paragraph above, you need to create a segment named "Term," and the use the NER feature to extract the dates. The first date will be the lease commencement date, and the second will be the lease expiration date.

Parsing of US and German addresses

Parsing models have been pretrained to extract US and German addresses. The models will recognize the following address components:

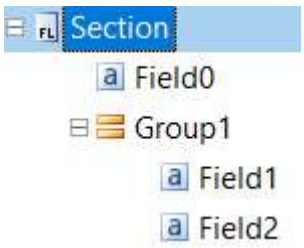
- Street (Line 1)
- City
- State (for US addresses)
- ZIP (postal code)
- Country

A text that you feed to the model may contain only one address. You can use the NER feature to detect and parse the address, or use a FlexiLayout to detect the address.

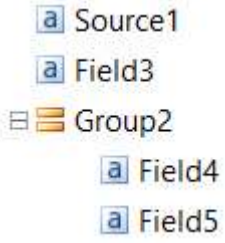


3.5. Limitations

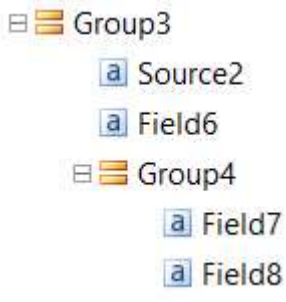
1. You cannot create multiple segmentation models within one document section.
2. Fields used in training cannot be nested more than one level deep. You can train a field inside a group of fields, but fields inside a group that is part of another group cannot be trained. The nesting level of a field into which an entity should be extracted starts at the source field. Possible options:
    a. Source: Section. A flat list of repetitive and non-repetitive fields in groups of non-repetitive fields



    b. Source: Text field. A flat list of repetitive and non-repetitive fields in groups of non-repetitive fields.



    c. A repetitive segment cannot serve as a source for an extraction model or an extraction script. To support a repetitive source, create a repetitive group.
    Source: Text field in repetitive group. A flat list of repetitive and non-repetitive fields in groups of non-repetitive fields inside the same group.



3. Quick markup of segments on web stations is not supported.

If you use autotraining based on feedback from verification operators, we don't recommend using web verification stations or FlexiCapture Cloud.

If you do not use autotraining, do not show segments at the verification stage:

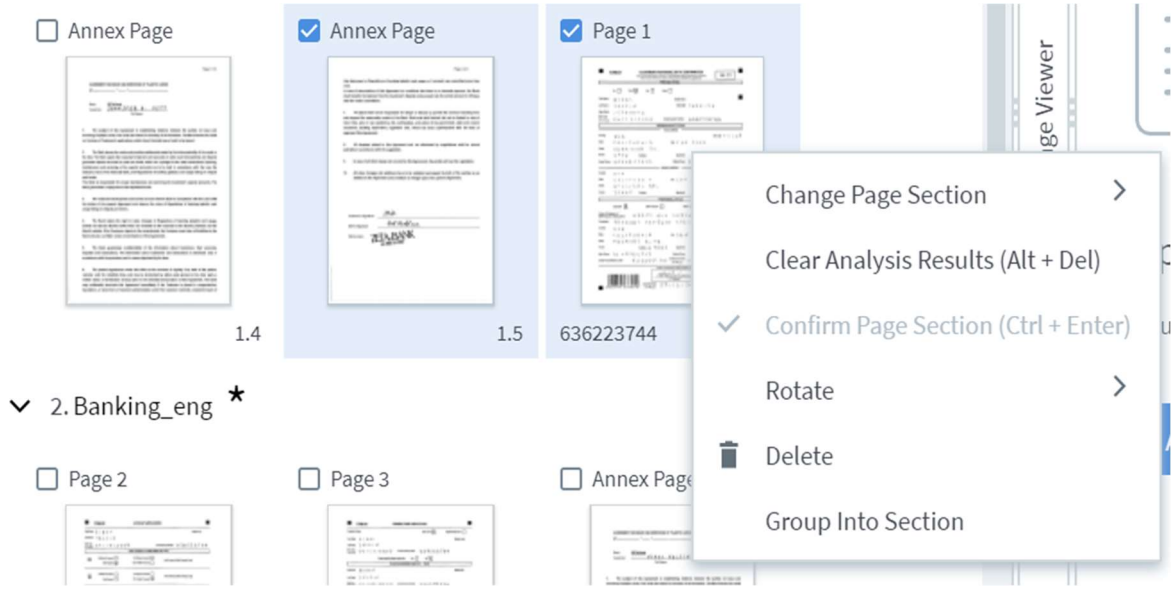
4. ABBYY FlexiCapture web applications

4.1. New command on web stations

Recognition of multi-page documents may pose certain problems. For example, one document may be incorrectly split into several documents, or an incorrect section may be chosen for some pages in a document (e.g. some pages may be wrongly identified as annex pages).

You can use the **Group Into Section** command to group several pages of a multi-page document into one section.

To do this, move the pages into one document, make sure that the right section has been chosen, right-click the pages, and then click the **Group Into Section** command on the shortcut menu.



4.2. Security improvements for web stations

ABBYY FlexiCapture 12 Release 3 Update 2 includes the following security improvements for its web stations:

- The CSRF issue was resolved for those end-points where it still persisted.
- Secure attribute has been added for cookies.
- Non-secure redirects have been eliminated.
- In the Remote Administration Console, a vulnerability has been eliminated that potentially allowed users to access other users' tenants.

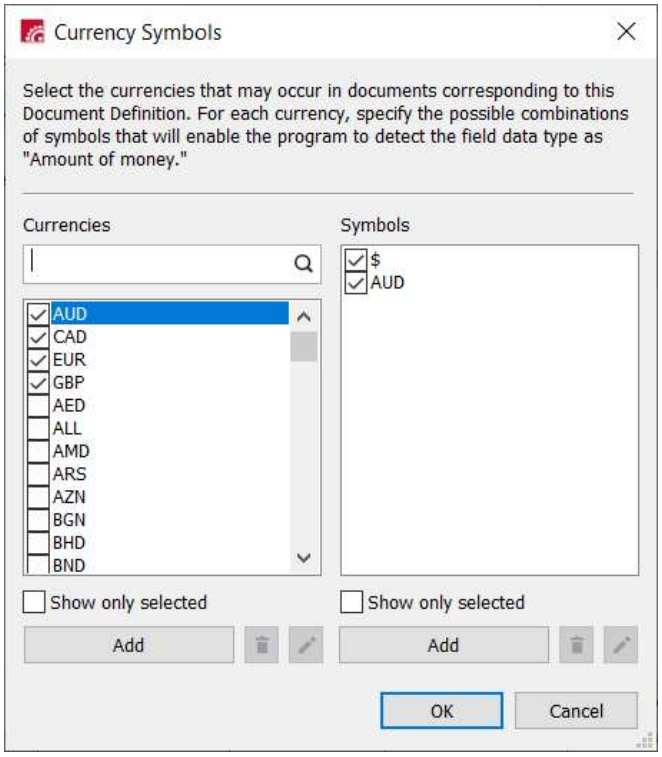
5. General improvements

5.1. New settings for "Amount of money" data type

When processing financial documents, it is important that amounts of money be extracted accurately. The data type used for extracting amounts of money allows letters, as these may occur in currency names. The downside of this was that letters were sometimes misread as numbers (e.g. O as 0, or l as 1). To avoid this from happening, you can now specify allowed currency names. The program will also know that currency names may occur before or after amounts of money, but not in the middle.

The allowed currencies are specified in the properties of a Document Definition and will apply to entire documents, i.e. to all the fields whose data type is set to "Amount of money."

To specify allowed currencies, open the Document Definition properties, click the **Recognition** tab, and edit the list of currencies in the **Amount of money** section.



You can select any of the standard currency abbreviations or add custom abbreviations and symbols.

5.2. Improved extraction mode for PDFs with a text layer

A PDF document may contain both a text layer and raster images (e.g. scanned pictures or embedded text).

Previously, two alternatives were available when processing this type of PDF document: you could either recognize the document in its entirety, or only extract the text layer, ignoring any raster objects.

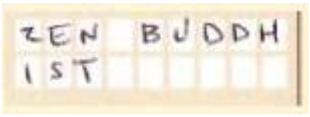
Starting from Update 2, ABBYY FlexiCapture 12 includes an "Auto" recognition mode, designed specifically for the type of PDF documents described above. In this mode, the program will take the text from the text layer, without performing any OCR, and recognize the raster objects. Should the text layer and a raster object overlap, the text layer will be preferred.

You can revert to the legacy recognition algorithm by modifying the following registry keys:

- For FlexiCapture: [Computer\HKEY_CURRENT_USER\Software\ABBYY\FlexiCapture\12.0\DAForms] EnablePullXTextAndRecognizeRest=true
- For FlexiLayout Studio: [Computer\HKEY_CURRENT_USER\Software\ABBYY\FlexiLayoutStudio\12.0\DAForms] EnablePullXTextAndRecognizeRest=true

5.3. Dictionaries now used for ICR fields with words divided at the ends of lines

Dictionaries can greatly improve recognition accuracy and are especially useful when capturing hand-printed text. Now dictionaries can also be used for words divided at the ends of lines (see example below):



Additionally, the **One line** option can now be used for fields with all types of marking.

All the above options are available on the **Recognition** tab of the **Document Definition Properties** dialog box.

5.4. New method for removing stamps

A new stamp removal method has been added that works especially well on CJK documents. As this method is rather slow (it may take up to several minutes to remove a stamp from one page), we recommend restricting the stamp area by passing only the rectangular region of the stamp to this method.

Method in scripts: IEditablePictureObject::RemoveStampsNN([in] IRects* _rects);

Method in Scanning Station scripts: IWorkspaceItem::RemoveStampsNN([in, optional] VARIANT/*IPageRect*/ _rect);

5.5. Circular navigation of verification objects

When the operator presses the F4 key (or clicks the corresponding menu command) to navigate the verification objects within a document, the program will sometimes jump to a verification object in the next document.

However, sometimes the operator may want to revisit the verification objects in the current document (e.g. to correct some rule errors which were skipped in the first pass).

On the desktop stations, you can now change the behaviour of the F4 key and its corresponding menu command by changing the value of the following registry key:

[HKEY_CURRENT_USER\Software\ABBYY\FlexiCapture\12.0\Shell\FormView\Behaviour]
"NextErrorInsideDocument"="True"

When set to “True,” pressing F4 (or clicking the corresponding menu command) will navigate through the verification objects only within the current document until all the objects have been verified (e.g. when all the low-confidence fields have been confirmed and all errors have been corrected).

The operator can also force the program to go to the next document by double-clicking a document in the list of documents or by placing focus on another document and pressing Enter.

5.6. Ability to handle large objects that do not fit on an A4 page

Some input files (e.g. XLS spreadsheets, TXT files, or e-mail messages) may contain very large objects that do not fit on one A4 page. This may be a table with a very large number of columns, a large picture in the body of an e-mail message, or a very long line in a TXT file. By default, the program will scale down large pictures and split large tables across

several pages, which results in Document Definition matching errors. To address the issue, we recommend reducing input files to A3 Landscape. The table below lists the registry keys that allow you to change the default page size.

	Option	Registry key	Key value
1	Scale inline images	Turned on by default	
2	HTML with table adjustment	HKEY_CURRENT_USER\ Software\ABBYY\ FlexiCapture\12.0\Controller\ OfficeConverter	PreprocessorOptions = 2 or PreprocessorOptions = 15
3	TXT with table adjustment	HKEY_CURRENT_USER\ Software\ABBYY\ FlexiCapture\12.0\Controller\ OfficeConverter	PreprocessorOptions = 1 or PreprocessorOptions = 15
4	Remove “thead” tag in HTML	HKEY_CURRENT_USER\ Software\ABBYY\ FlexiCapture\12.0\Controller\ OfficeConverter	PreprocessorOptions = 4 or PreprocessorOptions = 15
5	XML prolog removing for emails	HKEY_CURRENT_USER\ Software\ABBYY\ FlexiCapture\12.0\Controller\ OfficeConverter	PreprocessorOptions = 8 or PreprocessorOptions = 15
6	XLSX adjustment	HKEY_CURRENT_USER\ Software\ABBYY\ FlexiCapture\12.0\Controller\ OfficeConverter	AdaptPageSize = true
7	Autofit columns in XLSX	HKEY_CURRENT_USER\ Software\ABBYY\ FlexiCapture\12.0\Controller\ OfficeConverter	AutoFit = true

Known limitation: Scaling for TXT and HTML is not available if LibreOffice 4.2 or 4.3 is installed.

5.7. Changes to JSON export

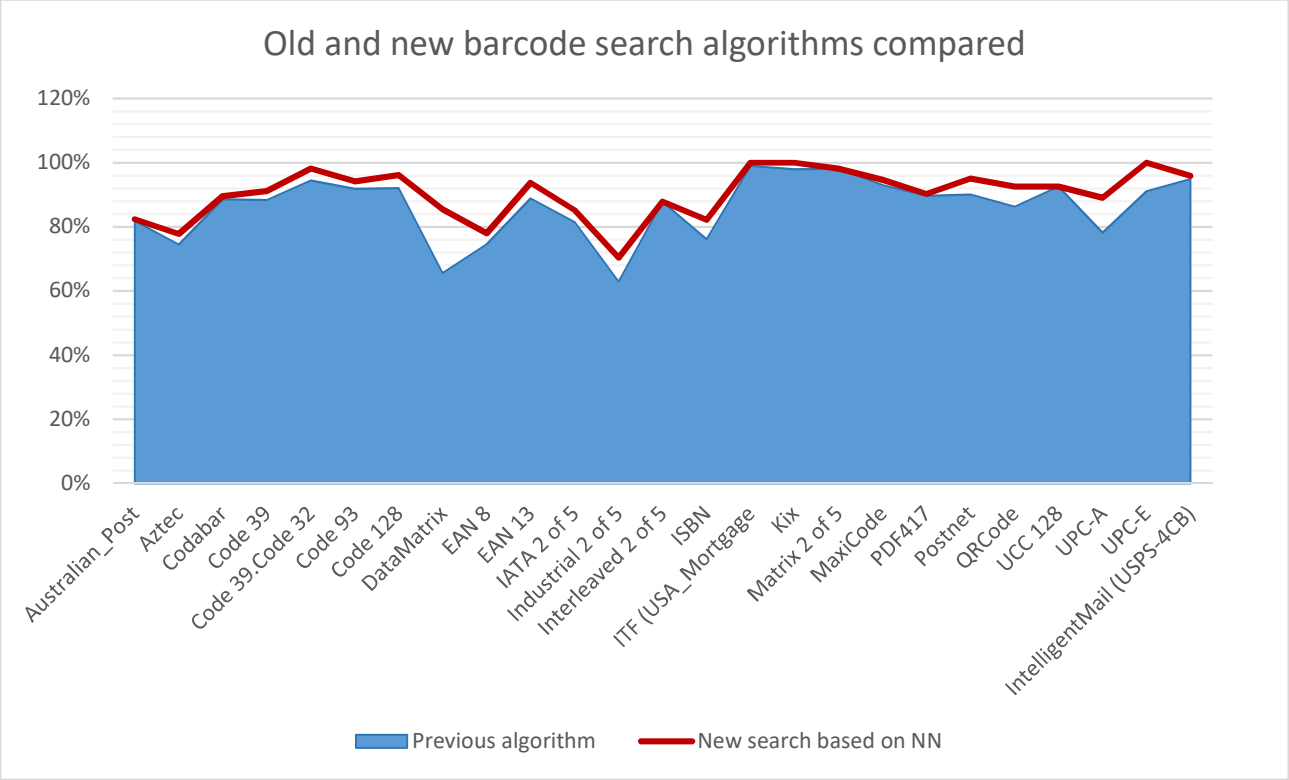
The following changes have been made to JSON export:

- Double values are now exported without brackets.
- JSON 1.1 is now used.

5.8. Neural networks are now used to detect barcodes

The old barcode search algorithm has been discarded in favour of neural networks. The new NN-based algorithm is more powerful and will detect barcodes in many complex cases, which could not be successfully handled by the old algorithm. For example, the program will now more easily detect barcodes positioned at a slant (a common occurrence when barcode stickers are affixed to documents).

The following diagram shows how the new NN-based algorithm compares against the old algorithm.



Our experiments have shown a slight increase in processing times (10-30 seconds per 100 documents) and slightly greater demands on RAM (the new algorithm requires an extra 100 MB of memory).

5.9. UI is now available in Italian

The UI of the desktop stations and the web stations is now available in Italian.

5.10. System Administrator's Guide is now available in German and French

The System Administrator's Guide is now available in German and French.

System Administrator's Guide in French
System Administrator's Guide in German

Bug Fixes

Issue description
Data set lookup worked incorrectly if the search string contained a space.
The number of reliably recognized characters displayed in details view changed after opening the Data Verification window.

Issue description
FlexiCapture for Invoices: Street was incorrectly displayed after a vendor search.
When exporting a project without batches, only the last Document Definition version should be exported.
Inline images larger than A4 were not scaled properly and could not be processed.
ORC-A type strings were not recognized correctly.
FlexiCapture for Invoices: The processing of Swiss invoices was affected by Russian or Ukrainian keywords.
Barcodes could not be detected on some images.
The Document Definition setting significantly slowed down verification.
FlexiCapture for Invoices: Negative numbers in the Total field were not recognized.
In some cases, the right borders of fixed-width fields were moved.
For XLS files containing columns with a lot of characters, the text in the cells was sometimes displayed as "#####" and could not be properly processed by FlexiCapture.
An error occurred when recognizing certain invoices.
Training results could be spoiled by one document in a training batch.
If an e-mail message attached to a batch contained certain characters in the Subject field, the program failed to export the project.
The Verification Station crashed when attempting to delete a large table with recognized line items.
New import tasks could not be created in any tenants after creating a test copy of a project.
Training failed for certain mark-up.
Default roles were sometimes deleted when editing project properties.
In some cases, the program failed to detect text located close to separators.
It was not possible to assign rights for a user if the user was imported from AD and the username contains special symbols.
The FullTextCharRects method did not return text.
An error occurred if an auto-replace script tried to replace any symbol on \r.
An error occurred during recognition of specific PDF files.
An error occurred during Document Definition matching.
Incorrect data was shown in the progress dialog during a text search.
PO Matching did not work for Web verification.
Repeating fields were not displayed on Web stations.
The project became invalid after upgrade.
FlexiLayout Studio crashed when the name of a column in a particular project was changed.
It was not possible to set the "skip training" flag for a document using scripts.
It was not possible to copy a project from the default tenant into another using FCAdminTools.
After being imported, batches got sent to Exceptions if the "Delete blank page" and "Separate Documents by Barcodes" settings were both enabled.
For the Standalone version, updating the DataSet did not work when using a script.
Processing failed when importing a cyclic AD group.
The Web Verification station did not close the session when switching to a different project.
An error occurred if a task was taken from the batch view and the RejectTask method got called.
An error occurred when importing a document from a second-level subfolder of the HotFolder.
A "Required registration parameter is empty" error occurred on the Web Scanning station even if a registration parameter was specified.
Opening the errors from the dialog window when closing a task caused the Verification station to crash.
Two operators got the same task.
An error occurred in some cases if EAN13 separation was used.
Using OpenDocument method caused an error in some cases.
A document could not be opened after it was placed into the Classification Training Batch when applying a splitting method.
Vendor look-up returning too many results slowed down recognition.
No page count when exporting data from the Batches page of the Administration and Monitoring console.
Batch list repeated 3 times when the Verification station was opened.

Issue description
Saving changes made to batch type settings slowed down performance.
A "Could not load file or assembly" error occurred when uploading a project to the server.
In some cases, a blank page opened when closing a task in the Web Verification station.
A session for a user with a custom role was not displayed in the Administration and Monitoring console until the user took a task.
Scanning station performance deteriorated when establishing a high-latency connection in a particular environment.
The Project Setup Station crashed when opening the Document Definition after upgrade particular project from FlexiCapture 11.
Margins between elements in a data form were modified when adding or deleting elements on the form.
An error occurred on the Project Setup Station when attempting to delete and add a checkmark region for a particular project.
Incorrect event order when using the On Open Document shell script.
Verification results were deleted when navigating away from the last field on the form during Web verification.
Web station sessions did not close for Google Chrome version 80.0.3987.116 and later and Firefox version 73.0.1 and later, which made the station run out of licenses, causing a "Cannot get license ticket" error message.
An error occurred when upgrading a database from FlexiCapture 10.
An error occurred when upgrading a project created in FlexiCapture 10.
Two Chinese hieroglyphs were not recognized correctly.
Disabling the CI_TextSizeLargest/Smallest commands had no effect.
Country-BU Check tax rates caused an error even if the specified value existed in the set.
An error occurred during pre-processing if a project was renamed on the server.
IBatch.Properties.Get() returned the old value for a modified registration parameter.
Deadlocks occurred during processing.
The "Recognized date" page field was updated each time a batch was opened.
Field values were only modified after a second attempt.
A large number of tables and repeating groups caused form mark-up to break down during verification on the Web station.
The digits 5 and 6 were mixed up during recognition.
The Update Project operation did not preserve Document Definition order.
An error occurred when trying to download a project with a training batch containing attached images that have long names.
An error occurred during the processing of a specific invoice.
Incorrect text structure when exporting to a *TXT file.
Upgrading a project from an old version caused an error to occur when uploading it to the server.
Text from a *PDF file with a text layer was not always extracted properly.
The SamlRequest URL was formed incorrectly if it contains Query Parameters. As a result, SSO with GSuite does not work.
Switching to the batch view from a batch on the Verification station did not preserve Id column sorting (the column header shows that the sorting is still applied).
Data columns containing auto-corrected symbols were not sent for group verification during the data verification process.
Data hidden using the “Redact sensitive data” option remained in the PDF text layer after exporting.
A Document Definition became invalid when upgrading a project from FlexiCapture 11.
Autoreplace did not work when changing the last symbol.
An error occurred when recognizing a particular document.
An error occurred when using the FullText method in IPage interface.
Old text was entered into the field that was being modified when working in “Automatic task receive” mode.
Some text was not recognized correctly on a particular document.
There was no diagnostic message when an incorrect class name was assigned on IPage::ResultClassName.

Issue description
An error occurred when trying to match an unstructured document definition using the IMatchingInfo.ForceMatch method.
If the “Check against database if completed” option was enabled, the “Enter value from database” option in the “Vendor must be found in database” rule stopped working.
In some cases, an error occurred when opening a project from a tenant for an Oracle database.
In some cases, no suggestions appeared when typing several symbols into a field.
The command to open the Classification panel was missing from the Main menu on the Verification station.
A “Referenced file not found: file_name” error when importing using SFTP (MOVEit SFTP).
FlexiLayout Studio pre/post search relations did not recognize External Fields of neighboring elements.
A field stayed highlighted as having an error even after HasRuleError was set to false.
Different sorting methods were used for looking for and processing images in the HotFolder, leading them to become incorrectly separated.
An inactive rule that referred to a non-existing method in the global rule module caused an error when opening a document on the Web station.
Typing error on a tooltip for the Close task button on the Web Station.
An error occurred when using IField.Type for the document set section.
Incorrect recognition results caused by the PC locale due to the "Writing Style" option being hidden.
Some line items were not extracted from images.
Large amounts of advanced pre-/post relation code slowed down FlexiLayout Studio performance.
For some documents, a class was not assigned using FCTools.ClassifyPage.
The Web Station UI language switched to English during page navigation.
Test Connection did not work on the Monitoring station when setting up an e-mail server for notifications.
An error occurred while updating DataSet using particular data.
A blank page opened when getting a task on the Web Verification station.
General operator report showed incorrect data in some cases.
If two processing stations both had a limit on allowed tasks, a registration parameter set during the execution of an automatic processing task on one station was not accessible from the other one.
Incorrect Tax Rate for Switzerland.
Processing of some PDFs failed with an error.
It was not possible to open a task on the Web Verification station if there was a field with language with extended language code (e.g. German - New Spelling).
The LoadColorPicturesByDefault key did not work in the Web station web.config file.
Recognition took too long for a particular document.
Total, Total Net и Total tax were not being detected correctly on some documents.
An error occurred when clicking on the Train button on the Project Setup Station.
An error occurred when processing documents created in StreamServe Communication Server.
Updating a project on the server failed due to the “Could not find the specified Document Definition or version” error.
After creating a new Au-NZ/CA Invoice project, the Invoice_old Document Definition was not removed from the definition list.
Classification in FlexiLayout Studio charged pages each time when executed.
Barcodes were not being recognized on some images.
The "I" character was not being recognized in FlexiLayout Studio.
The "-" character was not being recognized on some documents.
Using the DoneAnalyzePageObjects() method resulted in an error.
The UI became unresponsive when trying to load a classifier in classification training batch properties.
Some documents were being detected as invoices instead of credit notes.
Specific text objects were not being detected on some documents.
An error occurred when carrying out training using particular samples.

Update 2 Patch 1

Technical information

Release	Part #	Build #	Installer Build #	OCRT build #	Release date
Release 3 Update 2 Patch 1	1299/43	12.0.3.4040	176094	16.1.1014.15	2020.08.06

Bug Fixes

Issue description
Documents could not be recognized/analyzed after using Clear Analysis Result.
Data verification results were not being saved.

