

# ABBYY FlexiCapture 12 Release 1 Release Notes

---

# Table of Contents

---

INTRODUCTION .....	4
About This Document .....	4
About the Product .....	4
Installing ABBYY FlexiCapture 12 .....	4
Upgrading from previous versions .....	4
Licensing .....	4
TECHNICAL INFORMATION.....	5
KEY FEATURES .....	5
NEW FEATURES .....	5
1. Multitenancy.....	5
1.1. System Administrator and Tenant Administrator roles .....	5
1.2. Multitenancy licensing .....	7
1.3. Working without tenants.....	7
2. Service Level Agreement.....	7
2.1. SLA settings .....	8
2.2. SLA monitoring .....	8
2.3. Monitoring the SLA status on a Verification Station and on the Project Setup Station.....	8
2.4. Assigning extra-high priority to urgent batches .....	8
3. Training and Auto-Learning in ABBYY FlexiCapture.....	9
3.1. Training ABBYY FlexiCapture when setting up your project .....	9
3.2. Auto-learning.....	9
4. Training ABBYY FlexiCapture to detect field regions.....	9
4.1. Enabling training for existing FlexiLayouts .....	10
4.2. Training ABBYY FlexiCapture to detect field regions instead of creating a FlexiLayout.....	10
4.3. Training ABBYY FlexiCapture to detect field regions when setting up a project .....	10
4.4. Training ABBYY FlexiCapture to detect field regions if field locations vary greatly from document to document ..	10
4.5. Creating a FlexiLayout for a layout variant in FlexiLayout Studio .....	11
4.6. Training ABBYY FlexiCapture during documents processing.....	11
4.7. Training ABBYY FlexiCapture Standalone .....	11
5. Advanced classification.....	11
5.1. Classification profiles .....	12
5.2. Classification statistics .....	12
5.3. Precision/recall ratio.....	13
5.4. The training of the classifier by the administrator of the Project Setup Station .....	13
5.5. Creating an advanced classifier in FlexiLayout Studio .....	14
5.6. Importing a classifier developed in FlexiLayout Studio into FlexiCapture.....	17
5.7. Adding third-party classification technologies .....	18
5.8. Licensing.....	18
6. Web Stations using HTML5 .....	18
6.1. Web Scanning Station.....	18
<b>6.1.1. Synchronization of registration parameters .....</b>	<b>18</b>
6.2. Web Verification Station.....	19
<b>6.2.1. Integration into third-party systems .....</b>	<b>19</b>
7. Monitoring and reports .....	19
7.1. Using Microsoft Report Viewer to generate reports .....	19
7.2. Using Crystal Reports to generate custom reports.....	20

7.3.	Synchronization of users and groups with Active Directory .....	20
7.4.	Productivity pages and new summary .....	20
7.5.	Filtering monitoring data .....	21
8.	Document sets .....	21
8.1.	Creating Document Definitions for document sets .....	21
8.2.	Creating document sets on a Scanning Station .....	23
8.3.	Verifying document sets .....	23
8.4.	Working with document sets using the API .....	24
8.5.	Using scripts to assemble Document Sets .....	24
8.6.	Exporting document sets .....	25
9.	Processing digitally born documents .....	25
9.1.	Importing digitally born documents .....	25
9.2.	Processing digitally born documents .....	25
9.3.	High quality of data extraction for digitally born documents .....	26
10.	FlexiLayout Studio .....	26
10.1.	Column filter in the Batch window .....	26
10.2.	Classification and FlexiLayout project licensing .....	26
10.3.	Operations on multiple elements in the tree of elements .....	26
10.1.	Backing up FlexiLayout and classifier projects in FlexiLayout Studio .....	26
11.	Scanning Station .....	27
11.1.	Possibility to change parameters during re-scanning .....	27
11.2.	Cutting images .....	27
11.3.	Multiple crop .....	27
12.	Image preprocessing .....	27
12.1.	Using scripts to remove stamps .....	27
13.	Export .....	28
14.	Documentation .....	28
14.1.	Online help .....	28
15.	Miscellaneous .....	28
15.1.	Native support for 64-bit operating systems .....	28
15.2.	.Net Framework 4.x for scripts .....	29
15.3.	Skipping analysis and recognition when matching a document section .....	29
15.4.	Working with repeating fields and groups .....	29
15.5.	Using additional FlexiLayouts for a Document Definition with multiple .....	29

## Introduction

### About This Document

This document describes the features that have been implemented in ABBYY FlexiCapture 12 Release 1.

### About the Product

ABBYY FlexiCapture 12 is a new version of the FlexiCapture product whose main goal is to provide a stable and scalable Data Capture platform which can serve as a basis for creating regional and vertical solutions.

### Installing ABBYY FlexiCapture 12

ABBYY FlexiCapture 12 can be installed on the same computer as ABBYY FlexiCapture 10 and 11.

**Important!** ABBYY FlexiCapture 12 does not work with Microsoft SQL Server 2005 Express. Please install Microsoft SQL Server 2008 Express or later.

Please refer to the [System Administrator's Guide](#) for ABBYY FlexiCapture 12 for installation details.

### Upgrading from previous versions

ABBYY FlexiCapture 12 can be installed on the same computer where ABBYY FlexiCapture 10 or 11 is already installed. You cannot use the setup wizard to upgrade your ABBYY FlexiCapture 10 or 11 to version 12.

You can use ABBYY FlexiCapture and ABBYY FlexiLayout projects and 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 and ABBYY FlexiLayout project created in an earlier version of the program, it will be converted to the ABBYY FlexiCapture 12 format. Once a project is converted, it can no longer be opened in the earlier versions.

A detailed description of the upgrade procedure can be found in the [System Administrator's Guide](#).

### Licensing

To work with ABBYY FlexiCapture 12 Release 1, you need a serial number generated for this version. Serial numbers for previous versions of ABBYY FlexiCapture cannot be used with FlexiCapture 12.

The following new licensing options have been added:

- **Virtual machine usage.** Depending on selected options, the user can be allowed to install and run all components on a virtual machine or install and run stations only on virtual machines. Special permissions must be granted for working via Azure or Amazon.
- **Advanced classification.** If this option is enabled in a license, the user can train auto classification both in FlexiLayout Studio and in FlexiCapture and use it for classifying pages. For details see the "[Advanced classification](#)" section.

## Technical Information

Release	Part #	Build #	OCRT build #	Release date
Release 1	1299/12	12.0.1.263	15.0.1.233	2017.05.19

## Key Features

1. Multitenancy support
2. Service Level Agreement support
3. Auto-training of field extraction and classification
4. Advanced classification
5. Web Scanning and Web Verification Stations using HTML5
6. Improvements to the Administration and Monitoring Console
7. Document sets
8. Processing of digitally born documents
9. Improvements to FlexiLayout Studio
10. New export settings
11. Online Help

## New Features

### 1. Multitenancy

A multitenant system provides easy and secure administration across multiple workgroups and ensures independent workspace and data isolation for each entity. Multitenancy support facilitates using ABBYY FlexiCapture in large corporations with multiple business units or departments and makes it possible to provide services based on ABBYY FlexiCapture.

ABBYY FlexiCapture 12 allows creating several independent tenants. Projects, batches, users, and groups can belong to one tenant and be invisible and inaccessible from the others.

#### 1.1. System Administrator and Tenant Administrator roles

The **System Administrator** sets up and maintains the entire system, ensuring its stable operation and high productivity. The System Administrator only installs and sets up ABBYY FlexiCapture and monitors its productivity, he does not have access to the users' information inside the tenants.

A **Tenant Administrator** is responsible for document processing inside his own space called *tenant*. He can manage users' rights and monitor document processing inside his own tenant but has no access to other tenants' data.

System Administrator	Tenant Administrator
<ul style="list-style-type: none"><li>• Creates and manages database and file storage</li><li>• Creates and deletes tenants</li><li>• Monitors system productivity</li><li>• Can also assume the role of Tenant Administrator for the default tenant, so in the case of a non-multitenant system he can be the only administrator, as in previous versions of ABBYY FlexiCapture.</li></ul>	<ul style="list-style-type: none"><li>• Adds users and groups and manages their rights inside the tenant</li><li>• Creates and uploads projects to the tenant</li><li>• Monitors document processing inside the tenant</li></ul>

The administrator that creates a database will be automatically granted the System Administrator rights.

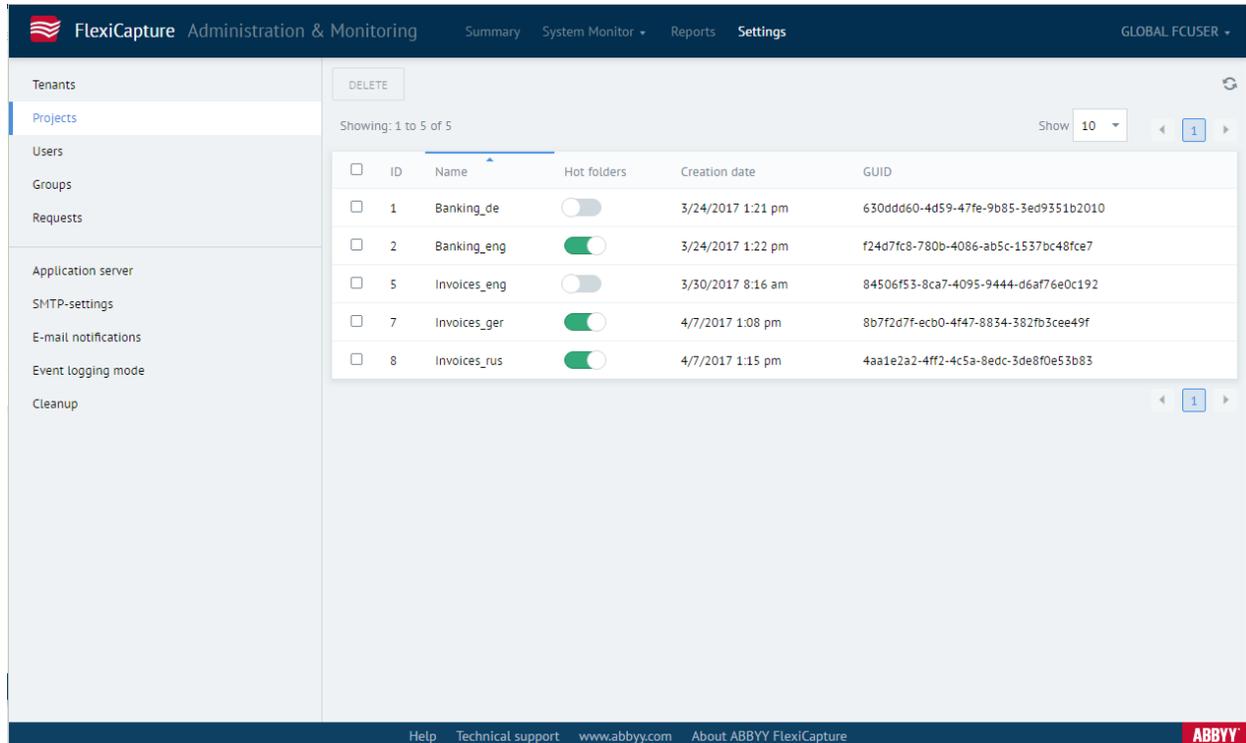
To add a new tenant, the System Administrator should open the Administration and Monitoring Console, select **Settings > Tenants**, and then click **New tenant**.

When creating a tenant, the System Administrator must enter the e-mail address of the person who will be the Tenant Administrator, and an automatically generated message with login and password will be sent to the indicated e-mail address.

Upon receiving the message, the Tenant Administrator can log in to the tenant using the URL <https://<server>/FlexiCapture12/Login/<tenant name>>. On first login, the Tenant Administrator will be asked to change his temporary password.

**Note:** Before creating a tenant, the System Administrator had to specify an SMTP server on the page **Settings > SMTP-settings**, so that an automatically generated message with login and password could be sent to the Tenant Administrator.

A Tenant Administrator can upload projects to the server within his tenant using the Project Setup Station, where he should select **File > Upload Project to Server...** and enter the URL of his tenant in the following format <https://server#tenantname>. In the window that opens, the Tenant Administrator must select the **Use server authentication** option and enter his Tenant Administrator credentials in the correspondent fields. To view the list of projects from the **Administration and Monitoring Console**, the Tenant Administrator must select **Settings > Projects**. The Tenant Administrator can allow the use of hot folders by a project from the list. The System Administrator can do the same for the default project.



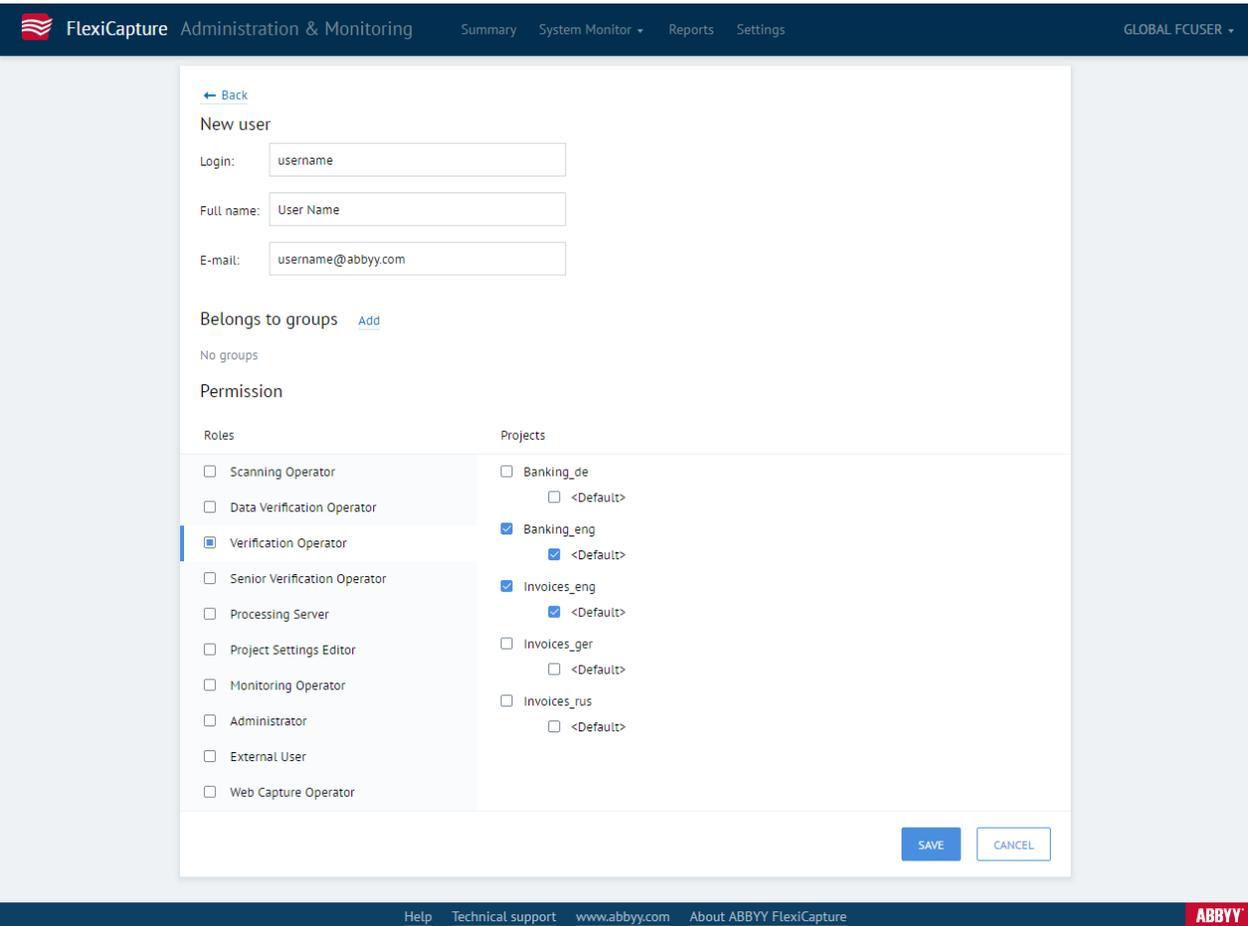
The screenshot shows the FlexiCapture Administration & Monitoring console. The left sidebar contains navigation options: Tenants, Projects (selected), Users, Groups, Requests, Application server, SMTP-settings, E-mail notifications, Event logging mode, and Cleanup. The main area displays a table of projects with columns for ID, Name, Hot folders, Creation date, and GUID. A 'DELETE' button is visible at the top left of the table area. The table shows 5 projects, with the first one being 'Banking\_de' and the last one 'Invoices\_rus'. The 'Hot folders' column contains toggle switches, some of which are turned on.

ID	Name	Hot folders	Creation date	GUID
1	Banking_de	<input type="checkbox"/>	3/24/2017 1:21 pm	630ddd60-4d59-47fe-9b85-3ed9351b2010
2	Banking_eng	<input checked="" type="checkbox"/>	3/24/2017 1:22 pm	f24d7fc8-780b-4086-ab5c-1537bc48fce7
5	Invoices_eng	<input type="checkbox"/>	3/30/2017 8:16 am	84506f53-8ca7-4095-9444-d6af76e0c192
7	Invoices_ger	<input checked="" type="checkbox"/>	4/7/2017 1:08 pm	8b7f2d7f-ecb0-4f47-8834-382fb3cee49f
8	Invoices_rus	<input checked="" type="checkbox"/>	4/7/2017 1:15 pm	4aa1e2a2-4ff2-4c5a-8edc-3de8f0e53b83

**Notes:**

- Please note that Hot Folders cannot be enabled from **Processing Server Monitor** as in previous versions of FlexiCapture. This functionality has been moved to the **Administration and Monitoring Console**.
- For automation scenarios, this functionality is also available in the **Processing Server Command Line Console**.

A Tenant Administrator can add users to the tenant by selecting **Settings > Users**. Operators can register themselves and send a request for rights to the administrator. If they have registered inside the tenant using the URL with the tenant name (see above), the request will be visible to the Tenant Administrator and he will be able to give access rights to a newly registered operator.



## 1.2. Multitenancy licensing

When creating a tenant, the System Administrator can allow a Tenant Administrator to use the main license or grant him a tenant license. The main license can be used by multiple tenants. However, this will result in certain limitations, such as a limit on the number of stations and a limit on the number of pages that can be processed.

By default, tenants use the main license to access ABBYY FlexiCapture. The **Share main license** option is enabled by default on the tenant's personal page in the Administration and Monitoring Console (**Settings > Tenants**).

The System Administrator can optimize the management of tenants by granting a tenant license to each tenant. In this case the System Administrator must activate a tenant's serial number in the ABBYY FlexiCapture 12 License Manager and set the serial number in the **License number** box on the tenant's personal page in the Administration and Monitoring Console (**Settings > Tenants**).

**Note:** If a tenant does not use the main license or a tenant license, the tenant will not be able to process documents.

## 1.3. Working without tenants

If no tenant is created, the default tenant will be used and the system will function identically to the previous version. The System Administrator will have the administrator rights for the default tenant and will be able to upload projects to the default tenant and give rights to users and groups.

## 2. Service Level Agreement

A lot of our customers have a Service Level Agreement with their end-users where certain document processing times are specified. Different document types might imply different timing. For instance, very important documents must be processed within a couple of hours, while conversion for archiving purposes does not require such tight deadlines.

ABBYY FlexiCapture 12 helps our customers to meet their SLA obligations, allowing them to specify expiration times for different batch types, and automatically prioritize and monitor batches that are about to expire. When a batch nears its expiration time, its priority is automatically raised, and the batch and its tasks are moved to the front of the queue. The

statuses of the batches and tasks are displayed on the Verification Station, on the Project Setup Stations, and in Administration and Monitoring Console, so users will be warned about problems in good time and will be able to resolve conflicts by allocating more resources to problem stages.

## 2.1. SLA settings

Allowed batch processing times can be set for each batch type on the **Workflow** tab of the **Batch Type Properties** and in the Batch Type Creation Wizard. The time starts when a batch is created on a Scanning Station (or on the server if no Scanning Station is used).

Select the option **Enable processing time limit for batches** and then click **Set Time Limit...** to adjust the settings. You can specify an absolute time period within which the batch should be processed, which will start once the documents are scanned (**Batch processing time limit**). You can also specify when a warning should be issued telling the user that he/she is running out of time (**Warn when \_ minutes remain**). If the setting **Issue warnings automatically** is selected, the warning time will be calculated by the server automatically, based on the average processing speed for the batch at the automatic stages. If the option **Set processing time limit with script** is selected, both absolute processing times and warning times can be customized by using a script. This can be useful in cases when batch expiration time depends on batch creation time (for example, batches created before 1 pm must be processed by 7 pm and batches created after 1 pm must be processed by 1 pm on the next day). Warning time can be also changed with a script. The time to start processing can also be read from a script.

- *SlaExpirationDate (DateTime, Read/Write)*: batch expiration time
- *SlaStartDate (DateTime, Read)*: time to start processing
- *SlaWarningPeriod (Int, Read/Write)*: warning time in minutes remaining until the batch expires

When creating a batch using the Web API, you can specify a batch creation time (*CreationDate*) and a time when processing should be started (*SLAStartDate*). Once a batch is created, the values of these fields cannot be changed.

If an expiration time is specified for a batch, the tasks in all the queues will be re-arranged in accordance with their expiration and warning times. As a result, the tasks closer to expiration will be selected first from among the tasks having the same priority.

## 2.2. SLA monitoring

If batch processing times are specified, you can monitor them in the Administration and Monitoring Console. The **Batches** view (**System (Processing) Monitor > Batches**) now contains default filters as: Expired, Expired soon. Clicking on Expired filter shows a list of expired batches, and clicking on Expired Soon filter shows a list of batches are close to expiration.

Each filter is editable and can be deleted. You can customize your own filter based on expiration data. To see all batches, which are already expired, you should set "Is Expired = Yes". To see all batches, which are close to expiration, set "Is Elapsed Expired = Yes" condition.

## 2.3. Monitoring the SLA status on a Verification Station and on the Project Setup Station

New columns – **Expires on** and **Status** – have been added to the following views:

Verification Station:

- Explore Queue
- Batches  
Using the batch filter in batch view, you can find all batches which have expired or are close to expiration.
- Postponed Tasks

Project Setup Station:

- Working Batches and Test Batches

The status **Warning** is displayed for batches that are about to expire and the status **Exceeded** is displayed for expired batches.

## 2.4. Assigning extra-high priority to urgent batches

Although task distribution is performed automatically to minimize SLA violation, sometimes it is necessary to raise a priority manually. The **Extra-High** priority setting allows pushing through urgent batches, documents, and tasks manually. This is the highest of all priorities.

Extra-high priority can be assigned manually:

- On a Verification Station, switch to batch view, right-click a batch, and from the context menu select **Change Batch Priority > Extra-High**. The selected batch will have extra-high priority until the batch reaches the **Processed** stage.

- On a Verification Station, switch to batch view, open a document, and from the context menu select **Change Document Priority > Extra-High**. When a document is given extra-high priority, the corresponding batch will have extra-high priority if it is to be processed in its entirety (this is specified in the stage settings). If the batch is divided into tasks, the tasks will be given extra-high priority.
- On a Verification Station, switch to queue view, click **Explore Queue**, right-click a task, select **Send Tasks to...** from the context menu, and in the **Send Tasks to Stage** dialog box change the task's priority to extra-high. The same can be done by opening a task (**Get task** command) and selecting **Mark document for sending to > Processing stage**. Extra-high priority assigned to a task will be valid for the selected stage only.
- Extra-high priority can also be assigned via the API. The TProcessingPriority object has a new value: PP\_ExtraHigh = 3. In the FlexiCapture database, extra-high priority corresponds to the value of "5."

### 3. Training and Auto-Learning in ABBYY FlexiCapture

You can train ABBYY FlexiCapture 12 to classify documents (i.e. identify document types) and detect field regions.

Page images and document text are used to identify the type of each document. ABBYY FlexiCapture 12 can easily cope with hundreds of different types of documents within one classification project.

You can train ABBYY FlexiCapture to detect the regions of individual fields and repeatable groups on documents with fixed and semi-structured layouts. If field locations vary greatly from document to document, you can classify documents into layout variants and train ABBYY FlexiCapture to detect field regions on each variant separately.

This allows you to:

- Set up your project quickly and start processing documents without creating FlexiLayouts.
- Add training functionality to any existing project, making the program learn from user input to improve the quality of document processing.

#### 3.1. Training ABBYY FlexiCapture when setting up your project

To set up a new project:

1. Define the document types that you are going to process. Create a Document Definition for each.
2. Collect some sample documents. We recommend having at least 10 sample documents for each document type.
3. Train the document classifier (for details, see "[The training of the classifier by the administrator of the Project Setup Station](#)").  
Note that you can improve the classifier manually (see "[Creating an advanced classifier in FlexiLayout Studio](#)").
4. Train ABBYY FlexiCapture to detect field regions on documents of each type (see "[Training ABBYY FlexiCapture to detect fields when setting up a project](#)").

#### 3.2. Auto-learning

You can make the program learn from user input to improve the detection of field regions (see "Improving the detection of field regions using operators' input").

ABBY FlexiCapture accumulates documents that have some field locations corrected by the operator during verification. These documents are copied to Training Batches after the Export stage and are then used for field detection training. Training starts automatically in the background, if training is allowed for the corresponding Document Definition Section and the Training Batch is not locked by the administrator.

The auto-learning feature ensures that the quality of field extraction doesn't deteriorate with the training and that system doesn't train indefinitely, accumulating an unreasonably large number of documents in the process.

Auto-learning is enabled by default. To disable it, go to the **Project Properties** or **Batch Type Properties** dialog box, click the **Workflow** tab, change the workflow schema to "Advanced," and disable the "Training" stage.

### 4. Training ABBYY FlexiCapture to detect field regions

You can train ABBYY FlexiCapture to detect the regions of individual fields and repeatable groups on documents with fixed and semi-structured layouts. If field locations vary greatly from document to document, you can classify documents into layout variants and train ABBYY FlexiCapture to detect field regions on each variant separately.

#### 4.1. Enabling training for existing FlexiLayouts

You can enable training in any Document Definition section that uses a FlexiLayout to detect field regions. To enable training, open the properties of the Document Definition section and select the **Allow field location training** option.

Note that training of fields locations is not allowed for fixed Document Definitions.

#### 4.2. Training ABBYY FlexiCapture to detect field regions instead of creating a FlexiLayout

You can skip the FlexiLayout creation stage and train fields from scratch. For this you need:

- Create a Document Definition, then create a section inside this Document Definition, and enable the **Allow field location training** option in the properties of the Document Definition section.
- Create document fields and select the **Can have region** option for the fields that you are going to extract from images.

Since there no field detection logic is available to the program at the very beginning, the operator has to indicate the location of each field manually. For this reason, we recommend enabling the **Store recognized text** option on the **Recognition** tab of the in **Document Definition Properties** dialog box. This will allow you to indicate field locations and fill in field values simply by clicking on fields.

Now you can start processing your documents and manually indicate field regions. The program will learn from this input and will soon be able to find the necessary fields automatically.

When you train the system, the FlexiLayout is created automatically. You can always access this FlexiLayout from the properties of the training batch, export it, and adjust it in FlexiLayout Studio.

#### 4.3. Training ABBYY FlexiCapture to detect field regions when setting up a project

1. Start the Project Setup Station.
2. Enable the **Allow field location training** option for the Document Definition section whose fields the program will be trained to detect.
3. Click **View -> Field Extraction Training Batches** and create a new batch.
4. A dialog box will open listing the Document Definition sections that are available for training. Choose the section whose fields you want to train. If you need to change the section, use the **Change Document Definition** command from the local menu of the batch.
5. A dialog box with a list of variants will open. Choose **Default Variant** for now (the choice of variant will be described later in the section "Training ABBYY FlexiCapture to detect field regions if field locations vary greatly from document to document").
6. Load document images into the batch and recognize them (or manually apply the appropriate Document Definition section).
7. Correct the locations of field regions where required. Note that you can only correct field locations but cannot correct field data. The program will be trained on those field regions whose locations have been corrected.
8. Be sure that documents are marked **For Training** (select documents and click **Set Document State -> For Training**). When you train the system, all documents with **For Training** label are used for training. Next, all documents that have either **For Training** or **For Testing** labels are used to assess the quality of the created FlexiLayouts and compare it with the previous version (if any) to prevent decrease in the quality of field extraction.
9. Train the program using **Project -> Layout Training -> Train Layout** command and open the log of the operation to review the result.
10. Now you can navigate to the working or testing batch and process some documents to see if the program has been trained successfully.

#### 4.4. Training ABBYY FlexiCapture to detect field regions if field locations vary greatly from document to document

While training works well for fixed forms and semi-structured documents, sometimes documents of one type may have very different layouts, termed *layout variants* in ABBYY FlexiCapture. For example, you may have to deal with variants within the document type "Tax Return," as different versions of tax return forms were issued over the years. Another example is invoices, where different vendors choose to position similar fields differently. To achieve the best training results for these kinds of documents, you need to train the classifier (see "[The training of the classifier by the Administrator of the Project Setup Station](#)") to differentiate variants and train ABBYY FlexiCapture to detect field regions separately for each variant.

To train ABBYY FlexiCapture to detect field regions on documents that come in different variants, you need to set up a list of possible variants:

1. Open the properties of the Document Definition section and click the **Data Sets** tab. The **Variants** data set stores a list of all layout variants for this the section. By default, it has only one record, "Default Variant."
2. Add a separate record for each variant.  
If you only have a few variants, you can add them manually. First, click the **View** button to open the list of variants and then click **Add** to add a new record. Alternatively, you can connect an external database as a source of variants (e.g, the list of suppliers) or even use a script to populate the list of variants taking them from a third-party system (click **Set Up** to use an external database of variants or variants from a third-party system).
3. Finally, save and publish your Document Definition.
4. On the Project Setup Station, open the **Classification Training Batches** view and train the classifier to differentiate variants (see "The training of the classifier by the administrator of the Project Setup Station"). If the classifier can't detect the variant of the document, the document will be handled as the default variant.
5. On the Project Setup Station, open the **Field Extraction Training Batches** view and train the program to detect field regions on each variant:
  - 5.1. Create a new batch.
  - 5.2. Select the Document Definition section whose fields the program will be trained to detect. If you need to change the section, use **Change Document Definition** command from local menu of the batch.
  - 5.3. Select the variant whose fields the program will be trained to detect. If you need to change the variant, use the **Change Variant** command from the local menu of the batch. Note that documents that you load in the batch associated with the variant should be classified as belonging to this variant, otherwise the program will not be able train or apply the FlexiLayout specific to this variant.
  - 5.4. Train ABBYY FlexiCapture to detect field regions as described in "Training ABBYY FlexiCapture to detect field regions when setting up a project."
6. After training the classifier to differentiate variants and after training the field detection algorithm for each variant, process some documents from the testing or working batch and review the results.

#### 4.5. Creating a FlexiLayout for a layout variant in FlexiLayout Studio

If you are not satisfied with the results obtained through training ABBYY FlexiCapture to detect field regions, you can create a FlexiLayout for a layout variant in FlexiLayout Studio. To speed up the process, you can export a trained FlexiLayout from FlexiCapture, adjust it in FlexiLayout Studio, and then import the adjusted FlexiLayout into FlexiCapture. (use the **Export Trained FlexiLayout** and **Import FlexiLayout** commands from the local menu of the training batch).

Note that a trained FlexiLayout contains all of the Document Definition fields, even though the program has been trained to detect the regions of only some of them. If you choose to adjust a trained FlexiLayout in FlexiLayout Studio, be sure to keep only the fields that should be detected by this FlexiLayout and delete the rest.

#### 4.6 Training ABBYY FlexiCapture during documents processing

If training is allowed for a Document Definition section, the program will automatically collect documents on which field locations have been corrected by operators and train itself on these documents. Self-learning will run as a background process. For details, see the "[Auto-learning](#)" section.

#### 4.7 Training ABBYY FlexiCapture Standalone

ABBYY FlexiCapture Standalone can be trained on sample documents just like ABBYY FlexiCapture Distributed. To train ABBYY FlexiCapture Standalone, the operator must select several document images and then click **Project > Training > Train**. The administrator mode of ABBYY FlexiCapture Standalone offers the same training capabilities to the administrator as ABBYY FlexiCapture Distributed.

## 5. Advanced classification

The classification technology included in ABBYY FlexiCapture allows customers to classify documents into different types (e.g. driver license, bank statement, tax form, contract, invoice, etc.) and their variations (e.g. invoices from different vendors) without applying sophisticated FlexiLayouts. It enables a mailroom scenario (i.e. automatic sorting of incoming documents) and simplifies the creation of FlexiLayouts by eliminating the need to define identification elements in a FlexiLayout, allowing to have separate FlexiLayouts for different document variations (e.g. separate FlexiLayouts for different vendors).

## 5.1. Classification profiles

Classification technology comprises

- image classification,
- text classification.

Image classification takes advantage of Convolutional Neural Networks. It distinguishes between documents that are different visually (e.g. invoices and contracts, invoices from different vendors, etc.). It works very fast — it takes about 0.1-0.3 seconds to classify one page.

Text classification relies on statistical and semantic text analysis. It requires OCR if a document was obtained by scanning, so in total it takes about 1 second to classify one page (the rest of the time is primarily OCR). However, this classifier provides extremely precise results on documents with a good text layer.

ABBYY FlexiCapture allows users to use any of these technologies separately or to take advantage of both of them simultaneously. The program can automatically combine the two technologies when necessary. There are two ways to combine image and text classification:

- **Fast** – The system learns for what classes image classification produces good results and doesn't use text classifier for those classes.
- **Thorough** – All the classifiers are applied for each page, and then an additional classifier combines their results and takes the final decision on the class of the page.

Image and text-based classification are termed "auto classification" because to train the classifier, you only need to provide a sample set with reference classes of documents and press the **Train** button. No rules or coding are required.

However, the developer of a classification solution can still adjust the result of auto classification using rules. The developer can use

- FlexiLayout-based technology, which can take into consideration words, text fragments, images, barcodes, separators, etc. and their position on the page (see Rules in the FlexiLayout Studio Classifier project),  
or
- scripts in .Net programming languages (see Classification Script in the Batch Type properties in the FlexiCapture project).

## 5.2. Classification statistics

Each document in a sample set has a reference class (or type) that is either assigned manually or at least verified by the user. During classification, the program assigns a class to each document. Classification statistics are an aggregate of the classification results for all the documents and describe how the classifier works on the provided sample set.

The statistics contain two groups of figures. The first group of figures is intended for users without special knowledge of document classification methods.

- **Correctly classified** – the number of documents whose assigned class is the same as the reference class provided by the user (including unknown documents classified as unknown).
- **Misclassified** – the number of documents that were classified incorrectly, so their assigned class is different from the reference class provided by the user (including unknown documents to which some class was erroneously assigned).
- **Not Classified** – the number of documents that were not recognized by the classifier, even though a reference class was provided for them.

The second group of figures is intended for users who are familiar with document classification methods.

- **Recall** – describes how many documents the classifier is able to recognize and classify correctly.
- **Precision** – shows how often the classifier is correct in its decisions and how often (100%-precision) it produces false positive results.
- **F-measure** – a value that describes overall classifier quality, combining both recall and precision with different weights.

**Note:** The program uses micro-averaging to calculate recall and precision across all classes, which means that a document class that is relatively rare in the sample set will contribute less to the classification statistics than a higher frequency class

with a lot of documents in the sample set. This method provides a realistic estimate of classifier quality when the number of documents of a particular class in the sample set is in proportion to what the user will encounter in real processing.

### 5.3. Precision/recall ratio

While 100% classification quality is hardly achievable in real life (even manual classification will produce from 1 to 3% errors on average), the user can set the “cost” for each type of error.

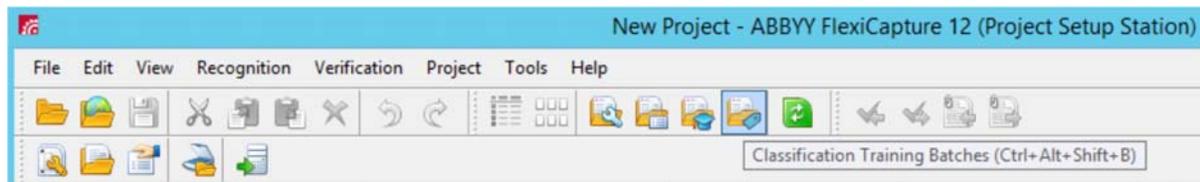
Let’s assume the user would like to have incoming invoices and legal contracts classified and forwarded to their respective departments. In this scenario, precision should be very high, i.e. the classifier should assign the correct class to each document, so that an invoice should never be forwarded to the legal department and a legal contract should never be forwarded to the accountants. At the same time, the fact that the classifier recalls only 90% of documents means that 10% of the document flow should be sorted manually.

Let’s consider another scenario, when a front end operator scans customers’ documents and the program automatically suggests document classes and performs integrity checks based on the classes known to it. In this case, the operator will go through almost all documents and it is better to maximize recall, i.e. the number of documents to which the classifier can assign the correct class, even if it will also slightly increase the number of false positive results, hence decreasing precision.

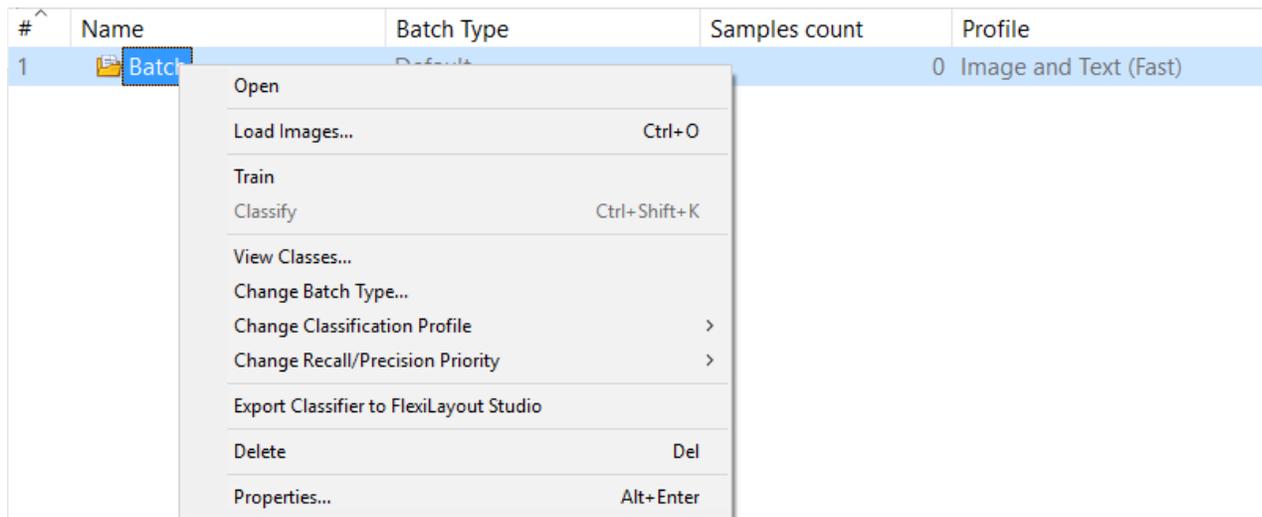
To accommodate for different real-life scenarios, ABBYY FlexiCapture allows setting the desired precision/recall ratio: the user can prioritize recall, precision, or use “balanced” mode.

### 5.4. The training of the classifier by the administrator of the Project Setup Station

Switch to the **Classification Training Batches** view and create a new batch.



Batch context menu:



You can choose a classification profile: image, text, or combined (Fast or Thorough). You can also prioritize precision or recall.

Load a sample set of documents into the batch (use %public%\ABBY\FlexiCapture\12.0\Samples\Images for demonstration purposes) and assign reference classes to the documents or pages using the **Set Reference Class** command on the local menu of each document/page.

If the project already contains Document Definitions with sections, each section can be used as a reference class. If sections contain variations, each variation may also be used as a separate class.

You can always create additional classes, but then you will need to map each new class to existing sections and variations. For example, if autclassification can't differentiate some sections, you can create a new class for those sections. In this case, classification will just narrow down the scope of available sections for matching to those sections. If a class corresponds to no sections, then the pages will be recognized as Annex pages.

Note that classification works for pages, but ABBYY FlexiCapture uses page classification results to makes decision on document types and on how pages should be assembled into documents.

The program needs about 10 pages of each class for successful training, though technically you can train the classifier starting with 1 document for each class. Using the **State** field, you can split the sample set into

- a sample set used for training and testing (select "For Training" state)
- a sample set used only for testing (select "For Testing" state)

Apply the **Train** command and see the log for the classification statistics.

**Note:** If you see the error "There are no pages to train the Classifier for the 'Default' Batch Type." make sure that you have enough pages in the "For Training" state.

To apply the classifier to real-life documents, you must link it with a batch type using the **Change Batch Type** command from context menu.

### 5.5. Creating an advanced classifier in FlexiLayout Studio

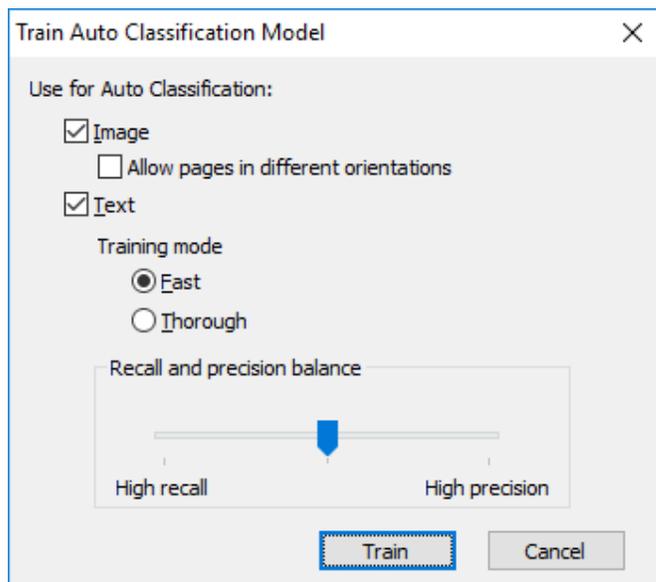
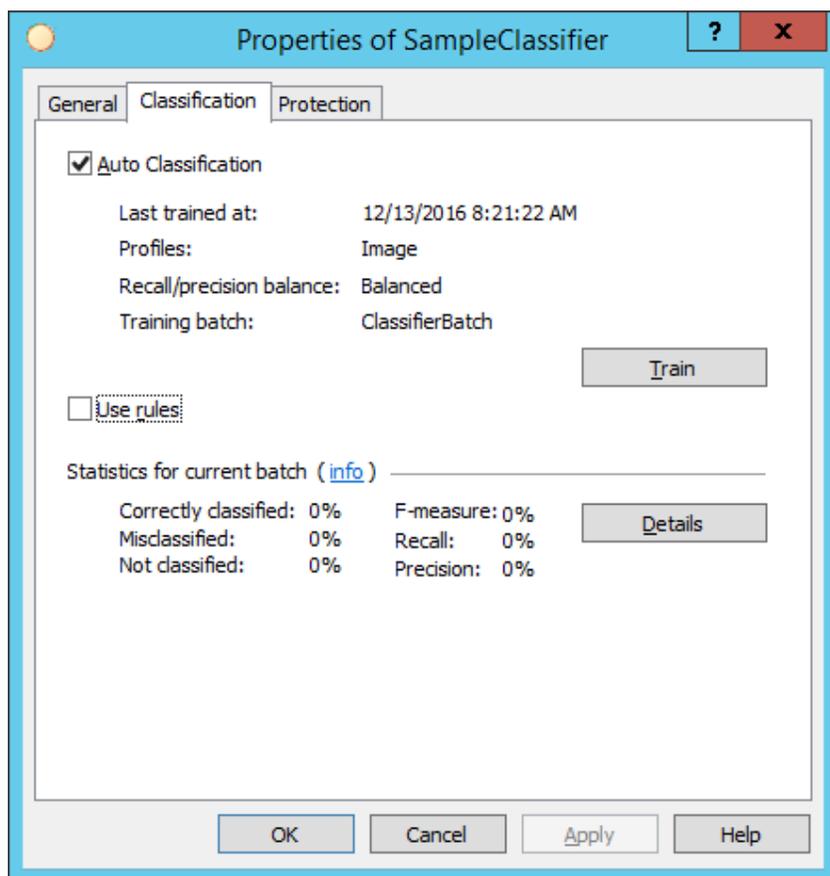
FlexiLayout Studio provides advanced tools for creating a production-quality classifier.

You can export a trained classifier from FlexiCapture Classification Training Batch by **Export Classifier to FlexiLayout Studio** command from the context menu the batch or create a new classifier project in FlexiLayout Studio. For demonstration purposes, you can use the standard sample project that is shipped with the product and is located in %public%\ABBYY\FlexiCapture\12.0\Samples\FLS\English\Classifiers\Auto&DT.

To create a classifier, the program needs a list of classes and a sample set of pages with reference classes assigned to each page. Typically, we recommend to have 100-1000 pages per class. FlexiLayout Studio allows working with multiple batches (sample sets) within one project, and is optimized to work with up to 10,000 pages per batch. If the sample set is huge, there are two ways to assign reference classes efficiently:

- You can create reference classes according to the source folder of each file (use the **Create Class by Source Folder** command)
- You can set classes manually for some of the documents, train the autclassifier, classify the rest of the sample set (or another portion), verify the assigned classes, and use the command **Set Classification Result as Reference Class**.

When the sample set is ready, you should train classifier. The Classification tab of Classifier Properties dialog box allows you to setup if you are going to use auto classification by image or by text, specify desired precision/recall balance and indicate if rules should be applied after auto classification. Also, this dialog box shows the result of classification for current batch.



You can specify a classification profile and prioritize precision or recall.

In addition to the settings available in FlexiCapture, you can use the **Allow pages in different orientations** option for an image classification profile. If this option is enabled, the technology will be trained to recognize the page regardless of its orientation (this may be useful, for example, for passports that can appear in different orientations within one document stream).

FlexiLayout Studio provides advanced statistics on classification results:

- Confusion Matrix, which visualizes how confusing the classes are and how many confusions happened
- Confusing Classes – the list of all cases of confusion, which allows you to sort out the most severe cases of confusion and resolve confusing classes one by one

- Statistics by Class, which allow you to identify low-quality classes

Current statistics (info) F-measure: 96% Recall: 96% Precision: 96%

	BankingFi	Contracts	Invoice	Letters	Prices	Questionnaire	BankingA	<None>
BankingFom	13							
Contracts	1	10						
Invoice	1		18					
Letters				10				
Prices					10			
Questionnaire						1		
BankingAgreement	1						10	
<None>								0

Confusion Matrix Confusing Classes Statistics By Class

Right-clicking a cell in the Confusion Matrix (or an item in other tools) will open a list of pages with confusing classes, allowing you to analyze specific confusions and come up with a resolution.

There are several ways to resolve issues if the autotclassifier does not work well for some documents:

- Enrich the sample set (a bigger sample set may include more variations to train the classifier).
- Merge classes that can't be differentiated by the image/text-based classification technology. Suppose image-base classification is used and the user has defined some visually distinguishable classes, but also included the full text of documents of different classes, e.g. "Contact A," "Contract B." These two classes cannot be differentiated by the classifier and this can reduce overall classification quality. At the same time, as a result of classification, some full-text document may already be of value for further processing, so the elimination of the two confusing classes "Contact A" and "Contract B" by merging them will improve classification quality.
- Create additional rules to differentiate classes.

To apply rules, select the **Classification by Rules** option in the classifier properties.

Rules are based on FlexiLayout technology. A rule is a tree of search elements that the program should or should not find in order to decide which class to assign to a document. A rule can decide if a document belongs to a particular class or not – this class is termed **Corresponding Class** in the rule properties.

To optimize the speed of classification, the user can specify the following for each class in its properties:

- whether rules should be applied if this class is assigned as a result of automatic classification
- whether all rules should be applied or only rules that can make a decision about this specific class (i.e. confirm or reject it)

Two types of classification errors may occur:

- 1) The page of reference class A is classified as Unknown. If you enable **Classification by Rules**, it will run for all pages that are classified as Unknown. So, you need to add a rule with corresponding class A. This rule should identify the page of class A among all the possible pages that can be classified as Unknown.
- 2) The page of reference class A is classified as class B (false positive). In this case you need to allow rules to run when the classification result is B (enable **Improve Auto Classification Result by Rule tree** in the properties of class B). Next you need to create two rules: one to decide that the page is A, and another for B.

If you don't want to create a sophisticated rule to retrieve all pages of a particular class but only want to differentiate A from B, you can create a differentiating rule by adding a condition in the **Pre-search relations** field of the obligatory element ID inside the rule:

```
// Get the auto classification result for the current document
```

```

StringArray AutoClassificationResult =
GetNamedParameterStringArray("AutoClassificationResult");
// Apply this rule tree only if auto classification result contains class
"SomeContract"
if not AutoClassificationResult.Has("B") then DontFind;

```

This rule will be applied only if the result of auto classification is B (or involves B).

You can extend the result of auto classification if some classes are often confused, e.g. you can add class A to B in the properties of class B. This will mean that if auto classification produces result B, then the program will automatically add A, and the auto classification result will be multiple classes: A, B.

Using rules to improve classification results for some class, you can specify in the class properties if all rules or only rules of corresponding classes should be applied.

The product team is working hard to make rules easier to understand and to create. In upcoming versions, the need to write code will be kept to a minimum, launch conditions will be added to the GUI, and some rules will be created automatically.

### 5.6. Importing a classifier developed in FlexiLayout Studio into FlexiCapture

A classifier works with pages and classes in FlexiLayout Studio, there is no idea of documents and Document Definitions there. When you import a classifier into FlexiCapture, you need to establish correspondences between classes and Document Definition sections and section variants.

To be clear, a class is the result of classifying a page. A section is a part of a Document Definition that should be matched to some set of pages. A class should correspond to one or several sections. If a section has variants, the class may correspond to one or several sections variants. It can correspond to the first page or any page of the section (by default, it corresponds to any page of the section).

When FlexiCapture processes a multipage document, it classifies the first page. The class defines the section (or list of sections) that should be matched. If the section has a FlexiLayout with an obligatory footer, then classification of other pages is skipped until the footer is matched. Otherwise, each next page will be classified and the end of the document will be defined either by the FlexiLayout or according to the classification result. To better understand how separation based on classification results works, consider the following example, which is a sequence of pages with classification results:

Page 1: Section A#AnyPage

Page 2: Unknown

Page 3: Section A#AnyPage

Page 4: Section A#FirstPage

Page 5: Section B#Any Page

This will result in documents: {Page 1, Page 2, Page 3}, {Page 4}, {Page 5}.

To export a classifier from FlexiLayout Studio, use **File -> Export Classifier**. This will produce a \*.cfl file that contains the classifier. To import it into a FlexiCapture project, open the **Project Properties** or **Batch Type Properties** dialog box and click the **Recognition** tab.

Correspondence between the classes from FlexiLayout Studio and the sections in FlexiCapture will be established automatically if class names follow this naming convention: Section[.VariantID][#"AnyPage"|"FirstPage"], e.g. "Section A", "Section A.4," "Section A#FirstPage," and "Section A.4#FirstPage" are valid examples of class names that will be automatically linked to sections and their variations. Section name is obligatory. If VariantID is not specified, then it will be linked to the default variant. If page mode is not specified, "AnyPage" is used.

If you want to establish correspondence between classes and the sections manually, switch to the **Classification Training Batches** view and select the batch associated with the batch type where the classifier has been loaded. Select the **View Classes** command from the local menu of the batch. In the dialog box that opens, you will see a list that combines the names of the section from the project's Document Definitions and the classes from the FlexiLayout classifier. Using this dialog box, you can link each class to one or several sections and variants and specify the first or any page of the section class.

### 5.7. Adding third-party classification technologies

FlexiCapture offers a classification script that allows users to combine the built-in classification technology with third-party technologies: cascading classifiers, voting, etc.

The classification script can be used to run the built-in FlexiCapture classifier and obtain a classification result for a document.

```
IBatchTypeClassifier result = BatchTypeClassifier.ClassifyPage(Page);
```

Now “result.ClassName” contains the name of the class assigned to the document by the classifier. The user can call any other classification engine, providing the Page object as input. When the final classification result is obtained, it should be written to the Page.ResultClassName property. The example below assigns to the page the class obtained from the FlexiCapture classifier.

```
Page.ResultClassName = result.ClassName;
```

### 5.8. Licensing

To use advanced classification, you need the corresponding permission enabled in your license. The basic license allows you to create a classifier project in FlexiLayout Studio and use only rules to create a classifier. If the **Advanced Classification** option is enabled in your license, you will be able to train auto classification both in FlexiLayout Studio and in FlexiCapture and use it for classifying pages. Page counter is decremented each time you classify a page in auto classification mode either in FlexiLayout Studio or in FlexiCapture. One page will not be counted twice. Any further operations performed on a classified page in FlexiCapture will not decrement the page counter.

## 6. Web Stations using HTML5

The FlexiCapture HTML5-based Web Stations ensure cross-browser compatibility and accessibility and offer a responsive and modern-looking design. No matter what sort of workstations are used — local, remote or web-based — companies with offices in multiple locations can easily implement distributed business processes anywhere they need.

There are two versions of Web Stations available in ABBYY FlexiCapture 12 — based on HTML5 and on Silverlight. The Web Stations based on HTML5 have reduced functionality compared to those based on Silverlight.

The new HTML5 versions can be opened by clicking the following links:

- Verification Station <http://<Server>/FlexiCapture12/VerificationNew/>
- Scanning Station <http://<Server>/FlexiCapture12/ScanningNew/>

### 6.1. Web Scanning Station

The HTML5-based Web Scanning Station currently offers the following functionality:

- No ActiveX or NPAPI component required
- Scan and load images and send batches to ABBYY FlexiCapture
- Imported batch types from ABBYY FlexiCapture on opening the station
- Rotate, flip, convert to b/w, despeckle, deskew, and invert images
- Create, merge, and delete documents
- Add batch and document registration parameters, fill registration parameters upon scanning
- Change scanning settings
- Get rescanning tasks from the rescanning queue
- Rescan pages
- Set batch type properties
- Zoom in/out of thumbnails
- Windows authentication support

#### 6.1.1. Synchronization of registration parameters

When an operator opens a project on a Web Scanning Station, the batch registration parameters are imported to the Web Scanning Station or synchronized between the server and the Web Scanning Station. Registration parameters created on the server cannot be deleted or modified on a Web Scanning Station; however, they can be deleted or modified on the server.

Modifications to batch scanning parameters will be synchronized between the server and a Web Scanning Station according to the following rules:

- A new parameter added on the server will become available on the Web Scanning Station.
- If a parameter is deleted on the server, it will also be deleted on the Web Scanning Station.
- If a parameter (e.g. "title," "required") is modified on the server, it will be updated accordingly on the Web Scanning Station.
- If a parameter (e.g. a default value) is modified on a Web Scanning Station, it will be preserved when updating the settings on the Web Scanning Station.

## 6.2. Web Verification Station

The HTML5-based Web Verification Station currently offers the following functionality:

- Select projects and roles
- Get tasks from a queue
- View batches, documents, tasks, and lists of postponed tasks
- Change priority of batches/documents
- Open the document editor and verify the main fields
- View all document errors
- Change Document Definitions
- Choose page section
- Merge documents
- Zoom in and zoom out of an image
- Select, create, and delete table regions on an image
- Select, drag-and-drop, and delete parent or child regions of repeating groups and fields on an image
- Create, edit, and delete table elements on an image (users can create, edit, and delete table separators)
- Bind table rows and columns on an image to table rows and columns on a data form
- Create, edit, and delete table regions and their elements (i.e. separators) on an image for multipage tables
- Select a vendor or business unit from a database using a lookup form
- View a list of errors and navigate through them
- Send document to exceptions/rescanning/stage
- Recognize document, check rules, apply changes
- Reject, postpone task
- Windows authentication support
- Integration into third-party systems

### 6.2.1. *Integration into third-party systems*

A Web Verification Station can be integrated into a third-party system via single sign-on technology (using SAML 2.0). Users will interact with the integrated Web Verification Station in the same manner as they interact with a non-integrated station. The web.config file can be modified to disable the header and footer of the Web Verification Station and customize the task menu.

*\* For more detailed information please refer to the accompanying documentation.*

## 7. Monitoring and reports

The new user interface ensures better business process continuity, visibility, and optimization. It provides comprehensive logging and reporting, multi-level administration, and automatic notifications for license expiration and new requests.

### 7.1. Using Microsoft Report Viewer to generate reports

Microsoft Report Viewer is now used as the default tool for generating reports. It is installed together with ABBYY FlexiCapture 12. Three standard reports are generated in Microsoft Report Viewer: site productivity, processing productivity, and general operators report.

Custom reports are also available in the Administration and Monitoring Console.

## 7.2. Using Crystal Reports to generate custom reports

Although Crystal Reports is not shipped with the product, users who have this product installed on their computers can continue using it for generating reports.

To use Crystal Reports, the following keys must be added into the Web.Config file:

```
<add key="UseCRRReportSystemForCustomReportOnly" value="true" /> – add this key to use Crystal Reports for generating custom reports (useful if a user has a lot of old custom reports which he does not want to re-design for Microsoft Report Builder). The key should be added to <appsettings></appsettings> section.
```

**Note:** Reports in Crystal Reports will be generated without a preview window and in PDF format.

If a user has custom reports in different formats, the correct software will be selected depending on the file extension: \*.rdl and \*.rdlc files will be opened using Microsoft Report Viewer and all other files will be opened using Crystal Reports.

## 7.3. Synchronization of users and groups with Active Directory

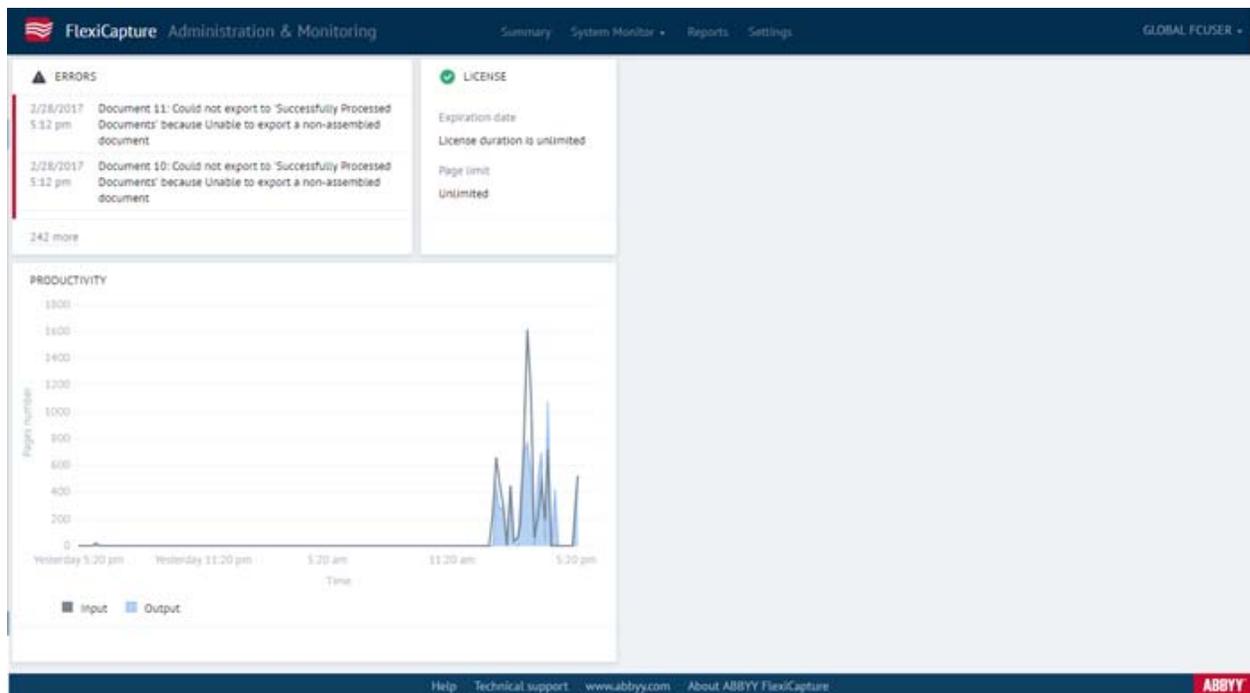
If users and groups are imported into ABBYY FlexiCapture from Active Directory, they can now be synchronized with Active Directory. This allows managing users and groups in Active Directory only.

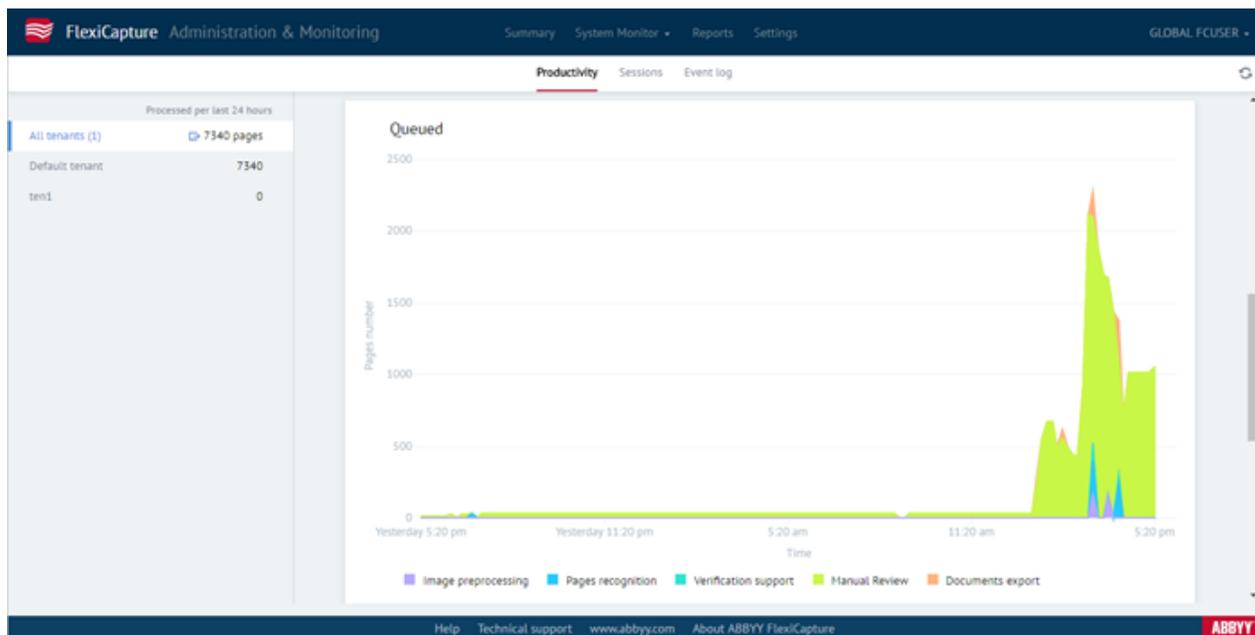
To synchronize users and groups, click the **Synchronize all groups** button on the **Import users or groups** page of the Administration and Monitoring Console.

A user added to an AD group will also be added to the corresponding ABBYY FlexiCapture group and will have all the rights assigned to that group. A user deleted from an AD group will be deleted from the corresponding ABBYY FlexiCapture group and the rights will be revoked. If a user is imported from Active Directory and is not included in any other groups, the user will be deleted from the ABBYY FlexiCapture database.

## 7.4. Productivity pages and new summary

The System Administrator can now check the system's productivity directly from the **Summary** page or using the page **System Monitor > Productivity**. The System Administrator can review overall productivity and productivity by tenants.





## 7.5. Filtering monitoring data

A flexible filtering mechanism has been added allowing you to display data in a convenient manner. Each monitoring view (Monitor > Batches/Tasks/Sessions/Event Log lists) has a set of predefined filters. For example, the Batches list has the Expired, Expires soon, and Exceptions filters. Clicking any of these filters will change the table view accordingly. Predefined filters can be edited or deleted. You can also create a new filter based on a predefined filter. Use the “+” icon to add new parameters and the AND and OR operators to link the parameters in your filter.

## 8. Document sets

A lot of our customers are processing documents that their clients submit to open a bank account, to get a loan, to get an insurance, etc. This case management scenario implies not only document recognition, but also a need to check that all the required documents have been provided and that all of them belong to the same case. ABBYY FlexiCapture 12 allows performing all these checks, displaying the main fields from different documents on the same form to enable comparisons of key fields, seals, photos or signatures, and displaying relevant data from the company’s databases to enable comparisons with extracted data. This functionality allows users both to extract the necessary data and make an informed decision about the case.

Please note that in Release 1 you can’t work with document sets on Web Stations.

### 8.1. Creating Document Definitions for document sets

A Document Definition for a document set can be created from the Document Definitions of the individual documents. A Project Setup Operator can create Document Definitions for passports, driving licenses, application forms, etc. and test them.

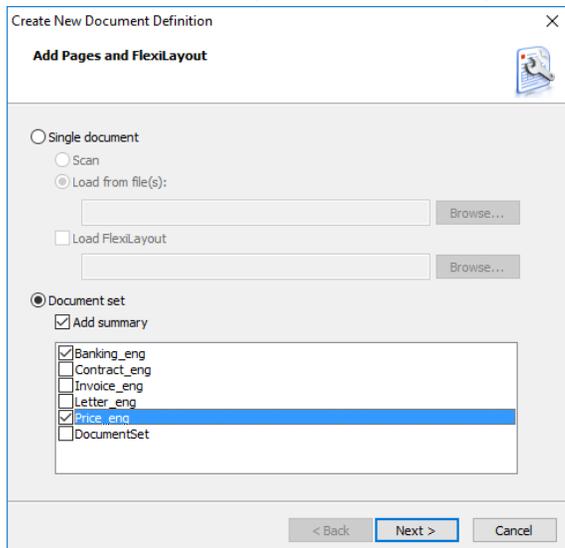
A Document Definition for a document set can have three types of section:

1. A **summary section**, which can be used to show the main fields from the child documents in the set. The summary section allows quick comparisons of fields (e.g. to check if all the documents have the same numbers or names) and shows additional information to the operator (e.g. additional data from the company’s database about the client who filled out the application form). The summary section does not include a page image, it only includes a data form.
2. A **document set’s own section**. It is identical to a single document section and can be created and edited as a section of a single document.
3. A **link to a Document Definition**. This section allows using one and the same document in multiple document sets (e.g. a passport can be a part of a mortgage application and an insurance case). A Project Setup Operator can update a Document Definition and the changes will be applied to all the document sets with a link to this Document Definition.

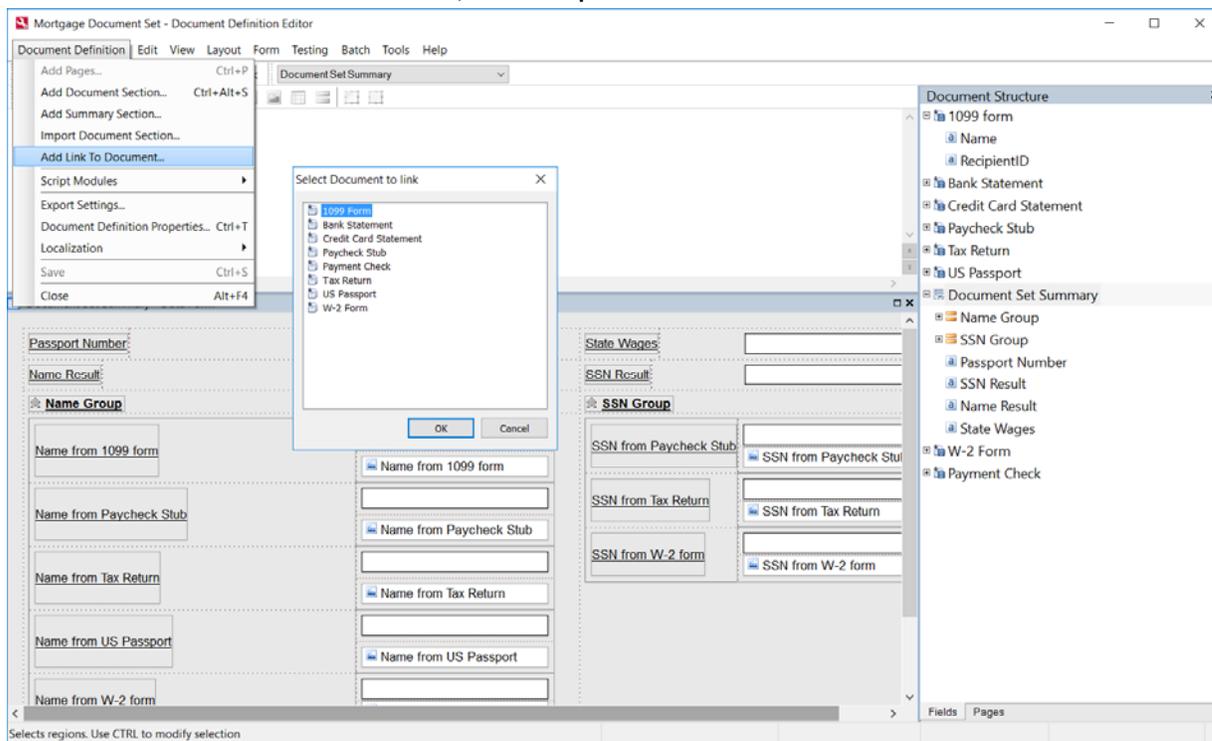
The above scenario can be implemented in one of the following ways.

On the Project Setup Station:

- There is a new wizard that allows creating a Document Definition for a single document or for a document set. In the case of a document set, you can add a summary section and links to existing Document Definitions:



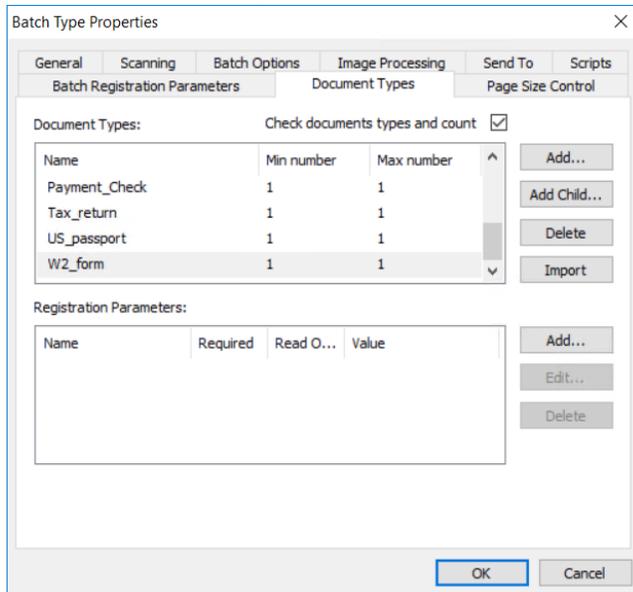
- It is possible to add links to documents and add a summary section using the **Add Link to Documents** and **Add Summary Section** commands in the Document Definition Editor.
- To create a document section from scratch, use the **Add Document Section** command. To add a document section created in another Document Definition, use the **Import Document Section** command.



- For each section except a summary section, you can specify the document set structure and the number of repetitions.
- The index fields are visible on a higher level, so if you need to display fields from child documents, or use them in rules at the document set level, or display them in the summary section, select the **Index field** option on the **General** tab of the field properties dialog box. After you add a new index field, use the **Update Index Fields** command from the context menu of the linked section. It is also possible to display an image of index field region by selecting the **Index field region** option on the **General** tab of the field properties dialog box.
- To place any fields in a summary section or in a document section, it is now possible to create **links to existing fields**. Such links can display values from child document fields and quickly open the respective child document in the document editor. To create a link, use the **Create Field > Link to Existing Field** command from the context menu of a section and then select a source document and an index field.

## 8.2. Creating document sets on a Scanning Station

- A Scanning Operator can manually assemble document sets after scanning. He can create sets and child documents and assign types to each of them. Document types can be imported from FlexiCapture server or created on the Scanning Station manually.
- To import document types, a Scanning Operator must click the **Import** button on the **Document Types** tab of **Batch Type Properties** dialog box. The imported document types will be displayed on the **Document Types** tab. Please note that the FlexiCapture server has to be specified to import document types. You can specify the server on the **Send To** tab of the same dialog box.



- If the contents of document sets must be checked before sending, select the **Check document types and count** option. All the document sets in the batch will be checked before they are sent to FlexiCapture and will not be sent unless their number matches the number specified in the settings.

## 8.3. Verifying document sets

On a Verification Station:

- Document sets can be assembled using separator pages or automatically when applying the Document Definition. Documents and document sets created in advance will be excluded from automatic assembly.
- During the verification operator can change documents assembly and move documents down and up within the hierarchy. A Verification Operator can verify and correct assembly in the thumbnail and document list views:
  - Using drag-and-drop
  - Using the **Split**, **Merge** and **Create Document** commands
- When the operator is sure about some documents boundaries he/she can manually fix a document's bounds, use the **Bound** command from the document's context menu. Such documents boundaries will be fixed and will not be re-assembled automatically.
- Documents included in a document set are now displayed as tabs to facilitate navigation.
- A Verification Operator can check the main fields in the summary section without opening documents or use the summary section to compare fields from different child documents. Links to the existing fields placed on the form facilitate navigation between documents (child document will be opened by click on the link to the existing field) and allow comparing key fields, seals, signatures, and photos. A message about errors in child documents is also displayed on the document set's form, so that the Verification Operator can quickly verify the entire document set.

1. Document Set Mortgage x 1.6. Paycheck Stub 1.7. Tax Return

1. Document Set Mortgage 1

Passport Number	453716397	State_Wages	16,680.24
Name_Result	HANY SAAD	SSN_Result	201-21-7421

**^ Name**

Name from 1099 form	HANY SAAD <b>Hany Saad</b>
Name from Credit Card Statement	HANY SAAD HANY SAAD
Name from Paycheck Stub	HANY SAAD HANY SAAD
Name from Tax Return	HANY SAAD
Name from US Passport	HANY SAAD
Name from W-2 form	HANY SAAD Hany Saad

**^ SSN**

SSN from Paycheck Stub	201-21-7421 201-21-7421
SSN from Tax Return	201-21-7421 201 21 7421
SSN from W-2 form	201-21-7421 201-21-7421

SSN, SSN, Employee\_SSN  
Compare SSN:Fields from compare rule do not match.

#### 8.4. Working with document sets using the API

- New fields have been added to the document structure: *ParentId*, the identifier of a parent document (which is 0 for top level documents) and *ChildrenId*, the identifier of a child document.
- *ChildrenOrder* is an array of type "Child ID – index for string."
- In the batch list, there is a document with id=0, which is a collection of top level documents. It initializes the order of child documents.

#### 8.5. Using scripts to assemble Document Sets

To meet the new requirements for Document Set assembly, the corresponding assembly script has been updated. The script has the following arguments: *BatchItems* is an array of child items of the current document and *AssemblingErrors* is an array of assembly errors.

If you need to upgrade a project created in an earlier version of the program, you can either edit and run your old script or delete the old one and start using the new script.

The following sample script generates a list of assembly errors and a list of pages and child documents.

```
string str = "";
foreach( IBatchItem item in BatchItems ) {
    if( item.Type == TBatchItemType.BIT_Document ) {
        str = str + "doc " + item.AsDocument.Id.ToString() + " " + item.AsDocument.TemplateName + " // ";
    } else if( item.Type == TBatchItemType.BIT_Page ) {
        str = str + "page " + item.AsPage.Id.ToString() + " " + item.AsPage.TemplateName + " // ";
    } else {
        str = str + "unknown type // ";
    }
}
AssemblingErrors.AddCustomError( str, 1 );
```

The following sample script gives you access to the index fields of child documents.

```
string namePassport = "";
string nameTaxReturn = "";
foreach( IBatchItem item in BatchItems ) {
    if( item.Type == TBatchItemType.BIT_Document ) {
        if( item.AsDocument.TemplateName == "US Passport" ) {
            namePassport = item.AsDocument.IndexedItemValue( "First and Last Name" ).ToString();
        }
    }
}
```

```

    }
    if( item.AsDocument.TemplateName == "Tax Return" ) {
        nameTaxReturn = item.AsDocument.IndexedItemValue( "First and Last Name" ).ToString();
    }
}
}
if( nameTaxReturn != namePassport ) {
    AssemblingErrors.AddCustomError( "Name from the passport (" + namePassport + ") and name from tax return form (" +
nameTaxReturn + ") does not coincide", 1 );
}
}

```

## 8.6. Exporting document sets

- Document set images can be exported to one PDF file or placed in a storage location.
- Document set fields (including fields in child documents) can be exported to ODBC databases and files.
- When a document set is exported, all child documents will be exported by the order they were added to the set. All fields in child documents can be exported. Setup export of fields from child documents in their respective Document Definition. From the Set's Definition you will have an access to the own set's fields and links to index fields of child documents.
- It is possible to redact sensitive information both in a document section and in linked documents.

## 9. Processing digitally born documents

ABBYY FlexiCapture smoothly processes office documents in all popular formats coming from any sources, including email attachments and message bodies. All documents can now be processed in the same flow with no need to separate image formats from text formats.

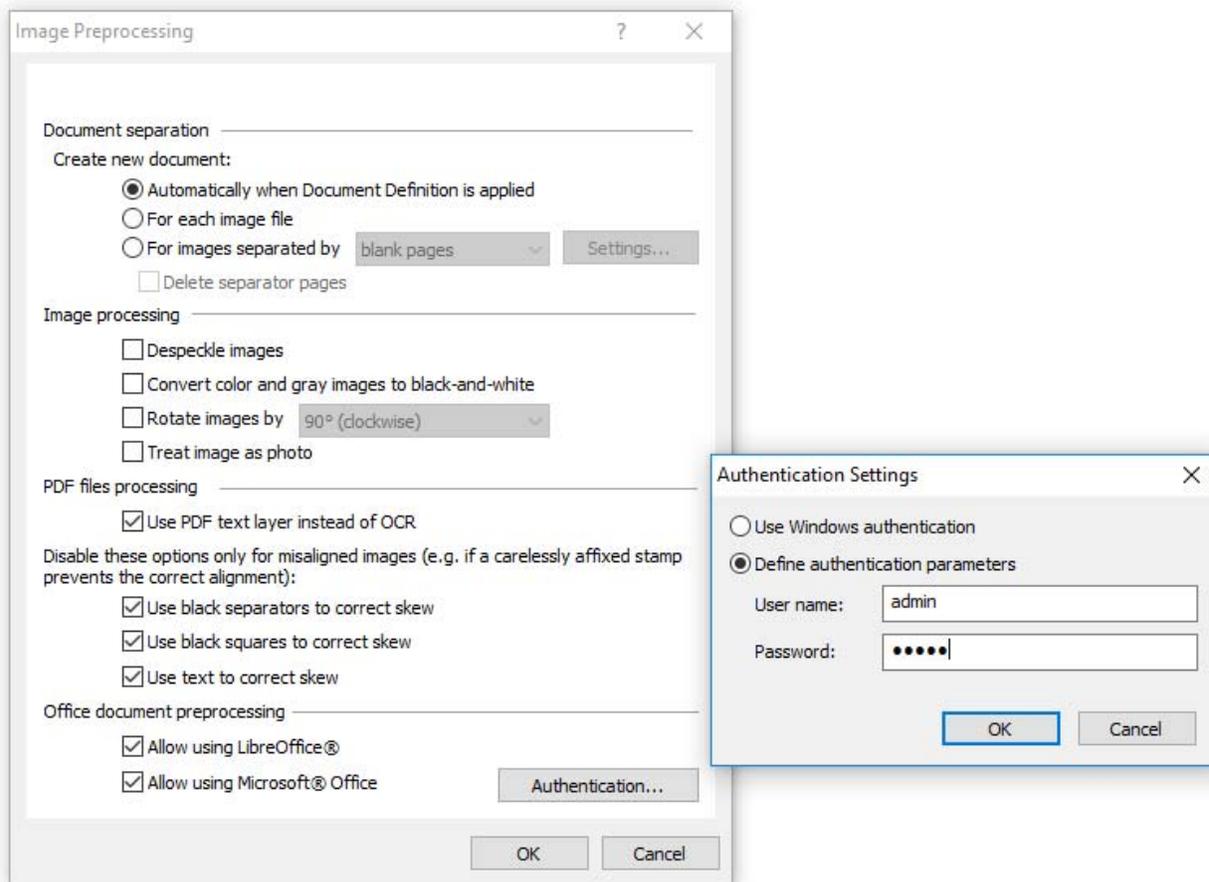
### 9.1. Importing digitally born documents

Files in following formats can now be imported into ABBYY FlexiCapture 12: TXT, EML, XLSX, HTML, DOCX, XLS, DOC, PPTX, HTM, PPT, and RTF.

### 9.2. Processing digitally born documents

One of the following tools can be used to process digitally born documents:

- Microsoft Office 2007 or higher if it is installed, its use is allowed in the settings, and a valid login and password have been provided
- Libre Office 4.0 or higher if it is installed and its use is allowed in the settings
- Apache Open Office built-in module for conversion if none of the above can be used
- To allow and set up the use of the software mentioned above, select the necessary options in the **Image Preprocessing** dialog box.



### 9.3. High quality of data extraction for digitally born documents

To achieve high quality of data extraction for digitally born documents, ABBYY FlexiCapture 12 always uses the text layer of Office documents obtained with Office software.

As for imported PDFs with a text layer, the text layer is compared to the recognized text and the text layer is used if it matches the text on the image. You can also enable option **Use PDF text layer instead of OCR** in case you are sure about original PDF quality.

## 10. FlexiLayout Studio

### 10.1. Column filter in the Batch window

An ability to filter documents by column has been added. To use the filter, right-click a column name and select the **Set filter** command from the context menu. A special filter sign appears next to the column name after the documents have been filtered.

### 10.2. Classification and FlexiLayout project licensing

It is possible to protect classification and FlexiLayout projects with a license. To protect a project, click the **Protection** tab and use the project context menu.

To use a protected project, a customer must have a license and the required module.

### 10.3. Operations on multiple elements in the tree of elements

Several elements can now be selected in the **Elements Tree** in ABBYY FlexiLayout Studio and commands will be applied to all of the selected elements.

### 10.1. Backing up FlexiLayout and classifier projects in FlexiLayout Studio

FlexiLayout and classifier projects can be backed up and restored. To make a backup copy, select the **Create Backup...** command from the **FlexiLayout** or **Classifier** main menu.

A **Backup manager** can be used to restore projects from a back-up copy, to delete back-up copies, and to view changes to the elements settings.

A back-up copy of a FlexiLayout project contains the tree of elements, the pre-recognition parameters, and the training elements. A back-up copy of a classifier project contains an auto classification etalon and rules, but no samples that were used for training.

## 11. Scanning Station

### 11.1. Possibility to change parameters during re-scanning

Batch type and registration parameters which are not read-only can now be edited at the re-scanning stage. It is also possible to change document assembly using drag&drop and context menu commands.

Please note that changing the batch type may result in the loss of the registration parameters and the results of previous processing. The Scanning Operator should change the batch type only if he/she is completely aware of the consequences of this operation.

### 11.2. Cutting images

In case documents of a small format were scanned on the same page, it is easy now to cut them into parts using new tool on the Scanning Station.

The tool allows cutting images using horizontal separator. Pressing shift button the separator becomes vertical, pressing ctrl button the tool can cut image in 4 parts using both horizontal and vertical separators at the same time.

### 11.3. Multiple crop

Another way to split small documents is by using the crop feature, which now allows selecting as many rectangles as you need. The final crop operation will create one document for each rectangle. You may prefer this method if you have a lot of empty space between your documents, as it allows you to split images and remove unnecessary margins in a single operation.

## 12. Image preprocessing

### 12.1. Using scripts to remove stamps

To support removing stamps from color images, new script methods have been added.

Methods added to IEditablePictureObject:

`NewRectsArray ()`: IRects – Creates a new empty array of regions

`RemoveStamps (color: TColorToFilter, rects: IRects)` – Removes stamps of the specified selected color from the specified region. Please note that this method can only be applied to color pages.

Methods added to IRects:

`AddRect (left: int, top: int, right: int, bottom: int)` – Adds a region to the collection.

New type TColorToFilter – selects stamps color to remove

Allowed color values:

```
CTF_Red = 0,  
CTF_Green = 1,  
CTF_Blue = 2,  
CTF_Yellow = 3
```

Methods added to IPictureObject:

`bool IsColor()`; – Checks if the image is a color image.

`bool IsGrayscale()`; – Checks if the image is a greyscale image. As the stamp removal method can only be applied to color images, it is recommended that you check if the image is color before applying the method.

The following sample script removes any blue stamps found in the entire page area if the image is a color image:

```
foreach (IPage page in Document.Pages)
{
    if (page.Picture.IsColor)
    {
        IEditablePictureObject editablePicture = page.Picture.CreateEditableCopy();
        IRects rects = editablePicture.NewRectsArray();
        rects.AddRect( 0, 0, editablePicture.Width + 100, editablePicture.Height + 100 );
        editablePicture.RemoveStamps( TColorToFilter.CTF_Blue, rects );
        page.ReplaceImage( editablePicture );
    }
}
```

## 13. Export

A new export setting has been added to create files for archiving in back-end systems with specific requirements for file formats, versions, and properties.

In the case of export to PDF, we offer a combination of settings that strikes a balance between file size and image quality. However, users can change the default settings to get files of minimal size or images of maximum quality.

- Users can now select a version of PDF/A. The list of available versions depends on the selected standard.
- The quality setting allows selecting between file size and image quality. Use “low” if you need the smallest possible files, use “high” if you need the best quality, or use “normal” for a balance between the two.
- MRC can now be used for non-searchable PDFs. Use this setting to get compressed PDFs that are small in size but of good visual quality.
- The “Fast web view” setting allows showing the first page of a multipage document before the entire document has been loaded. Use this setting for documents that will be put up on web pages. This setting is ignored for one-page documents.
- Use the “Enable tagged PDF” setting for searchable PDFs if you need to add tags.
- A compression codec can now be selected for color and grayscale images exported in TIFF. Select “Custom quality” and then select a desired compression algorithm.
- UTF encoding without BOM is now supported for all text formats (CSV, TXT, and XML).

## 14. Documentation

### 14.1. Online help

All information from the offline documentation is now available online. In the future, more help information will be available online than in the offline documentation. Help information is available in English and Russian. Online help information will be updated more frequently than the offline documentation. To access online help, visit <http://help.abbyy.com>.

Readers can leave feedback for any help item using a feedback form:

Please leave your feedback



The image shows a feedback form with the following elements:

- A label "Name" followed by a text input field.
- A label "E-mail" followed by a text input field.
- A label "Comment" followed by a larger text area with vertical scrollbars and a small icon on the right side.
- A red "Submit" button at the bottom.

## 15. Miscellaneous

### 15.1. Native support for 64-bit operating systems

ABBYY FlexiCapture now natively supports 64-bit operating systems. There are two installation packages, one for 32-bit systems and one for 64-bit systems. The appropriate package is selected automatically depending on the user’s operating system. Users can also install the required version manually by running the Setup.exe file from the corresponding folder in the ABBYY FlexiCapture distribution pack.

#### 15.2. .Net Framework 4.x for scripts

Now .Net Framework 4.x (4.0 and higher) can be used for creating scripts (rule scripts, export scripts, etc.) in ABBYY FlexiCapture. If .Net Framework 4.x is used, we recommend recompiling all scripts in .Net Framework 4.x and updating them in ABBYY FlexiCapture. Scanning Station also supports .Net Framework 4.x (4.0 and higher) for creating scripts.

#### 15.3. Skipping analysis and recognition when matching a document section

It is now possible to select a Document Definition for a selected document section without any analysis and recognition. This is useful if a section cannot be matched automatically (for example, the pages are of very poor quality and the data cannot be extracted). This also allows you to quickly select a Document Definition to check the assembly of a document or a set of documents, or in scenarios where the operator will fill out the fields manually.

To skip analysis and recognition, select a document section in the document list on the Verification Station or on the Project Setup Station, right-click it, select **Match Document Definition** on the shortcut menu, and switch off the option **Analyze and recognize after matching**.

#### 15.4. Working with repeating fields and groups

In the previous versions of the program, a repeating group could only be created in ABBYY FlexiLayout Studio. Now repeating fields and groups can be created in the Document Definition Editor, where a new property named **Can be repeated** is now available for text, barcode, and image fields and for groups. This property allows creating repeating fields and groups, and can be used for training or for manual filling.

#### 15.5. Using additional FlexiLayouts for a Document Definition with multiple

One Document Definition can now work with multiple FlexiLayouts. Among these, one is a main FlexiLayout and the other are additional FlexiLayouts. The main FlexiLayout is responsible for document identification and defining the document borders. The additional FlexiLayouts will only be used if the main FlexiLayout is matched. Additional FlexiLayouts can add more fields or re-define already matched fields. The following rules are used to combine fields from different FlexiLayouts: each subsequent FlexiLayout will either add a field with a unique name, or overwrite a field's region if a match was found in one of the previous FlexiLayouts. You can organize communication among FlexiLayouts within one Document Definition using named parameters.