

ABBYY FlexiCapture 8.0 Professional User's Guide

Dear User,

This guide explains the core principles behind ABBYY FlexiCapture 8.0 Professional. Please read this guide carefully before using the program.

For more information, please consult the following documentation:

- Online help can be accessed from the menu or by pressing F1. The help file can also be accessed from Start > Programs > ABBYY FlexiCapture 8.0 Professional > Helps.
- The ABBYY FlexiCapture 8.0 Professional System Administrator's Guide can be accessed from Start > Programs > ABBYY FlexiCapture 8.0 Professional > Guides > Administrator's guide.
- The operator's guide can be accessed from Start > Programs > ABBYY FlexiCapture 8.0 Professional > Guides > Operator's guide.
- The guide for creating machine readable forms can be accessed from Start > Programs > ABBYY FlexiCapture 8.0 Professional > Guides > Creating a Machine-Readable Form.
- Online help for FlexiLayout Studio 8.0 and FormDesigner 8.0 can be accessed from the menu of the corresponding application or by pressing F1. The help file can also be accessed from Start > Programs > ABBYY FlexiCapture 8.0 Professional > Helps.
- The ABBYY FlexiLayout Studio tutorials can be accessed from Start > Programs > ABBYY FlexiCapture 8.0 Professional > Guides > Tutorials.

To learn how to use the program effectively, we recommend that, in addition to reading the help files, you also set up a project for processing documents of a certain type. As an example, you can use a special *User Questionnaire*, a copy of which is located on the **<disk name>:\Documents and Settings\All Users\Application Data\ABBYY\FlexiCapture\ 8.0\Samples\FormDesigner\English\Questionnaire** (For Microsoft Windows Vista **<disk name>:\Users\Public\ABBYY\FlexiCapture8.0\Samples\FormDesigner\English\Questionnaire**). As you go through this guide, you will see instructions in boxes. These will teach you the main steps involved in program set-up and document processing, and allow you to view the results once the completed questionnaire has been processed. If you experience difficulties setting up the project, open a ready-made project called *Questionnaire.fcproj*, which can be found in **<disk name>:\Documents and Settings\All Users\Application Data\ABBYY\FlexiCapture\8.0\Samples\FlexiCapture\English\Questionnaire** (For Microsoft Windows Vista **<disk name>:\Users\Public\ABBYY\FlexiCapture8.0\Samples\FormDesigner\English\Questionnaire**).

This and other project samples can be accessed from Start > Programs > ABBYY FlexiCapture 8.0 Professional > Helps > FlexiCapture sample projects.

Please export the results to XML format and send them to us. Your feedback will help us learn more about our users and improve the product to better meet your requirements.

We trust that you will find our program pleasant to work with and enjoy using it.

Table of Contents

1.	Introduction	4
1.1.	The aim of document and data capture	4
1.2.	Automated document and data capture	4
2.	Administrator and Operator Functions	5
3.	What Documents Can Be Processed with ABBYY FlexiCapture 8.0 Professional?	5
4.	Setting Up the System for Capturing Fixed Forms	7
4.1.	Creating a fixed form template	7
4.1.1.	Form elements	8
4.1.2.	Marking of data fields	9
4.2.	Creating a project	11
4.3.	Creating a document template	12
4.3.1.	Document Template Editor	13
4.3.2.	Using elements to mark objects on the form	14
4.3.2.1.	Field groups	15
4.3.2.2.	Fields without marking	15
4.3.2.3.	Marking of Tables	16
4.3.2.4.	Fields with multiple regions	16
4.3.2.5.	Fields with multiple instances	17
4.3.2.6.	Excluding a region from recognition	17
4.3.2.7.	Deleting fields	18
4.3.3.	Static elements	18
4.3.3.1.	Peculiarities of barcodes	19
4.3.4.	Field properties	19
4.3.4.1.	Common field properties	19
4.3.4.2.	Data types	20
4.3.4.2.1.	Data types of a text entry field	21
4.3.4.2.2.	Data types for checkmarks	23
4.3.4.2.3.	Data types for a checkmark group	24
4.3.4.3.	Field recognition properties	25
4.3.4.3.1.	Recognition properties of a text entry field	25
4.3.4.3.2.	Recognition properties of checkmarks and checkmark groups	27
4.3.4.3.3.	Barcode recognition properties	29
4.3.4.3.4.	Picture recognition properties	29
4.3.4.4.	Verification parameters	29
4.3.4.5.	Picture export options	30
4.3.4.6.	Rule-based checks	31
4.3.5.	Creating a template for a multi-page document	33
4.3.6.	Creating a template for a document with annex pages	35
4.3.7.	Setting up data export	36
4.3.7.1.	Export to files	36
4.3.7.2.	Export to a database	37
4.3.7.3.	Export to a SharePoint library	39
4.3.7.4.	Custom export	41
4.3.8.	Setting up the recognized data view	42
4.3.9.	Editing and publishing a template	42
4.1.10.	Testing the template	42
4.4.	Setting up image import	43
5.	Setting Up the System for Processing Flexible Documents	44
6.	Peculiarities of Capturing Non-Structured Documents	45
7.	Working with a Set-Up Project	46
7.1.	Adding images	46
7.2.	Recognition	47
7.3.	Verification	48
7.4.	Export	52
8.	Conclusion	53

1. Introduction

1.1. *The aim of document and data capture*

All sorts of different documents are now used everywhere, by businesses, industries, and services alike. Applications, surveys, and invoices are an important part of the work of every enterprise or institution. The current standard of information technology makes it impossible to operate with paper documents only: most data must be converted into electronic format for storage, analysis, and further processing.

The main difficulty in processing a paper document is in entering the data into a computer system. Traditional manual data input can only be justified if the volume of information is relatively small. If the volume of information is large, manual input is inefficient because it is far too labor-intensive, slow, and expensive. Manual input cannot be made more efficient overnight and the time and expense incurred by an increase in manual input is more or less equivalent to the cost of starting the entire process cycle from scratch.

Manual data input is clearly far from ideal. The alternative is much less problematic and much more efficient – an automated data and document capture system. ABBYY FlexiCapture 8.0 Professional is exactly such a system.

1.2. *Automated document and data capture*

ABBYY FlexiCapture 8.0 Professional is a software system for capturing data from documents of various types: fixed forms, semi-structured documents, and non-structured documents.

Automated data capture involves the following steps:

- **Scanning** of a batch of pages using a scanner;
- Automatic assembly of scanned pages into documents;
- Automatic character **recognition**;
- If the program has generated multiple hypotheses about certain characters (i.e. the recognition of these characters is uncertain), the characters are sent to the **verification** operator;
- **Export** of verified data to a file or database and saving of document images to the specified folder. Images may be saved in graphical format or searchable PDF;

ABBYY FlexiCapture 8.0 Professional is an efficient solution for automated data capture that allows you to monitor and manage data processing and to control the quality of input data.

The benefits of using ABBYY FlexiCapture 8.0 Professional

- **Low cost.** The system can be set up so that most operations are performed automatically, almost completely without human intervention. The operator need only enter batches of documents into the scanner feeder and verify uncertain characters. Instead of several workplaces devoted to manual data input, you therefore only need one workplace equipped with ABBYY FlexiCapture 8.0 Professional.
- **Faster input.** Automating the entire process can significantly speed up document processing.
- **High recognition quality.** ABBYY award-winning technologies allow fast, high quality character recognition. It is extremely difficult to achieve comparable quality with manual data input without a significant drop in processing speed.
- **Convenient to use and easy to learn.** ABBYY FlexiCapture 8.0 Professional offers a user-friendly interface, both for the system administrator and document-processor. No lengthy

training courses are required to work with the program. Users can master the program easily with the help of our technical documentation and reference materials.

- **Scalability.** Unlike a manual input system, ABBYY FlexiCapture 8.0 Professional is easily scalable. To increase output, you need only install the system at an additional workplace.

2. Administrator and Operator Functions

To capture data using ABBYY FlexiCapture 8.0 Professional, you need to set up the system to process a certain document type. Administrators specify the necessary settings and operators capture the documents themselves. The number of workstations depends on the volume of documents to be processed.

Administrators specify the system settings and monitor the system. Their duties include:

- Creating new forms (this may also be undertaken by a designer);
- Setting up image import and scanning parameters;
- Preparing document templates, including:
 - Setting up recognition parameters;
 - Specifying verification and document assembly rules;
 - Setting up data export;
- Monitoring the process of data input in the set-up system, advising operators;

Operators are responsible for inputting documents. Using the system as set up by the administrator, they process completed documents of a certain type. Their duties include:

- Scanning and importing documents;
- Verifying data;
- Exporting data;

Document recognition is automatic.

The program has two modes: Administrator mode and Operator mode. The Administrator has full access to all program functions. Administrator mode can be password-protected. An Operator cannot access template creation and editing or import profile creation. An Operator can work with only one project or only one batch.

When you install the program you can choose between installing only the Operator Station or the complete program. If you select complete installation, you will be able to switch between administrator and operator mode. To select one of the modes, select either **Start > Programs > ABBYY FlexiCapture 8.0 Professional > ABBYY FlexiCapture 8.0 Professional Administrator Station** or **Start > Programs > ABBYY FlexiCapture 8.0 Professional > ABBYY FlexiCapture 8.0 Professional Operator Station**.

3. What Documents Can Be Processed with ABBYY FlexiCapture 8.0 Professional?

ABBYY FlexiCapture 8.0 Professional is a software system used to capture data from documents of various types.

Documents of various types can be processed in a single batch, and you can set up the system for processing documents of a mixed type. A document type only affects the method by which the document template is created. The type of processed documents does not affect the work of an operator.

Let us analyze the various document types that can be processed using ABBYY FlexiCapture 8.0 Professional.

- **Structured documents.** Structured documents are documents containing a certain number of specific data fields whose location and marking are identical on all copies of the document. Such documents are called “fixed forms”. Questionnaires, surveys, and application forms are usually fixed forms and are typically paper forms that must be filled out by hand. To identify a specific form from a flow of various documents and to extract data from such a form, you need to create a uniform fixed template that tells the program where to find the necessary data fields. Certain fixed forms are processed more efficiently because they were created to meet specific data capture requirements. Such forms are called “machine-readable forms”. ABBYY FormDesigner 8.0 is an efficient tool for creating machine-readable forms and is supplied together with ABBYY FlexiCapture 8.0 Professional. For more information on creating forms using ABBYY FormDesigner 8.0, please read the ABBYY FormDesigner 8.0 Help file and other documentation. The main steps for creating a template are described specifically for structured documents.
- **Semi-structured documents.** These are documents containing a number of data fields whose quantity, marking, and location may vary on different copies of the document. For example, invoices are semi-structured documents because invoices received from different companies often differ with regard to the number of data fields and their format. All invoices have an invoice number and a total payment amount, but these data fields may be located in different places on the document. To identify semi-structured documents and to extract data from them, ABBYY FlexiCapture 8.0 Professional uses flexible templates (FlexiLayouts). To create a flexible template, you must use the special ABBYY FlexiLayout Studio module. For detailed information about this module, please refer to the ABBYY FlexiLayout Studio Help file and User Guide. Flexible document processing differs from fixed document processing only with regard to the creation and attachment of a template.
- **Non-structured documents.** ABBYY FlexiCapture 8.0 Professional can be used to process non-structured documents such as contracts, letters, or orders, where the information is presented in a free-form style. The program can automatically identify non-structured documents as attachments to fixed or semi-structured documents, or with the help of a flexible template. These documents can then be exported to searchable PDF files or to graphical files. Data from the index fields of non-structured documents can be extracted manually, or automatically using a flexible template. A typical scenario in which non-structured documents are processed is the conversion of a paper archive into electronic format and the extraction of several index fields for subsequent quick attribute search.

The following chapters describe how to set up ABBYY FlexiCapture 8.0 Professional to process documents of various types, including the automated data capture process, how to improve recognition quality, and how to organize data export.

The capture of structured documents is described in great detail. All processing stages are described using fixed forms as an example. The peculiarities of other document types and the differences concerning the creation of templates for such documents are also explained.

4. Setting Up the System for Capturing Fixed Forms

ABBYY FlexiCapture 8.0 Professional allows you to capture and process fixed forms quickly and efficiently. The process of working with fixed forms is as follows:

- Designing forms or making existing forms machine-readable;
- Creating a template: describing the geometrical layout of objects, specifying object properties, creating verification and document assembly rules, setting up data export;
- Setting up the method for adding images and creating image import profiles;
- Document capture, which can start once all the necessary settings are in place.

4.1. Creating a fixed form template

The design of a paper form is very important because its appearance determines whether it is convenient for a user to fill out this form and if the form is suitable for automated data capture. It must be clear to the user where he or she is required to enter the necessary data, and such clarity significantly reduces the risk of errors. When designing a form, you must try to make it as intuitive as possible to fill out. The data entered should also be clearly discernable and easily recognizable.

If a form meets all data capture requirements it is *machine-readable*. Such forms have the following properties:

- They possess *anchors* (or *reference marks*). These are special auxiliary elements, which help the program determine the form orientation, match the template, and deskew the scanned image where necessary. Anchors can be represented by black squares, crosses, or corners.
- All fields and graphic elements (separators, anchors, etc.) must be located in exactly the same place on all copies of the form. All the forms of a given print batch must be created using a single master copy.
- All explanatory information is positioned so as not to hinder the extraction of information from data fields.

For more information about machine-readable form requirements, please see the ABBYY FlexiCapture 8.0 Professional Help file and the "Guide to Creating Machine-Readable Forms".

ABBYY FormDesigner 8.0 is a special tool for creating machine-readable forms and is supplied with ABBYY FlexiCapture 8.0 Professional. For information on creating machine-readable forms in ABBYY FormDesigner, please refer to the ABBYY FormDesigner Help file and User Guide. ABBYY FormDesigner allows you to create machine-readable forms simply and easily. A form template created in ABBYY FormDesigner can be imported into ABBYY FlexiCapture 8.0 Professional. You will then have a complete layout of all data fields and graphical elements and only need specify the properties of the fields and set up the export.

To make an existing form machine-readable, you need only make some slight design changes and add certain element (in particular anchors). If you cannot change the blank forms for any reason, you can set up the program so that it can process forms without anchors. You can use other form elements as anchors, for example, vertical and horizontal lines, explanatory text, or barcodes. Template matching will however be most efficient if you use standard anchors.

4.1.1. **Form elements**

Let us analyze the main *form elements*. (Figure 1)

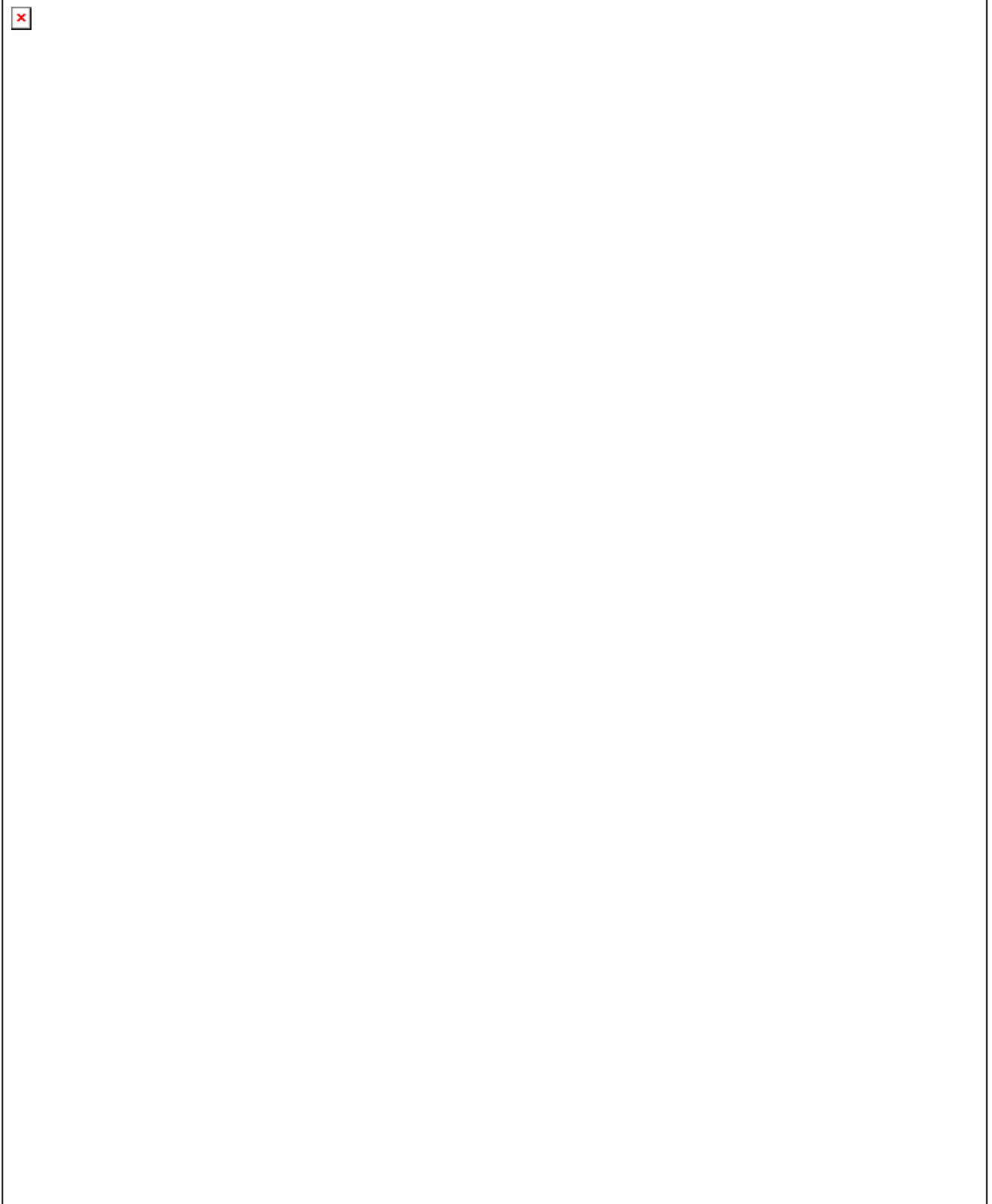


Figure 1. Sample of a machine-readable form containing the main elements

- **Data fields.** All forms designed for information gathering contain data fields. These fields are usually accompanied by an explanatory text. Data fields can be of the following type:

Text fields used to enter text information. Such fields are groups of character cells for entry of characters. The design of a text field prompts the person filling out the form to use separate characters.

Checkmarks are also designed for information gathering. However, users need not write any text but only mark the necessary items. A checkmark usually has a closed contour (a square, circle, or polygon), and the information is entered by entering a certain sign (for example, a tick or cross) inside the contour. Sometimes a checkmark does not have a contour, in which case the user must position the sign against the white background in the specified place on the page. If you wish to allow correcting checkmarks, select the corresponding option when creating the template. In such cases an inked-out checkmark will be considered unchecked.

Checkmark groups. A checkmark group consists of several checkmarks located close to each other and connected logically. As a rule, answers corresponding to checkmarks within a single group are mutually exclusive. In other words, only one checkmark in a group may be checked.

Data fields can also be represented by tables.

- **Anchors.** Anchors are used to determine page orientation and to match the template. The program also uses reference elements to monitor and correct (deskew) image distortions that may appear during scanning. Anchors can be represented by black squares (preferably), crosses, or corners. We recommend that you use 5 anchors on a form: four at each corner and one at one of the sides, in order for the page orientation to be reliably detected. Template matching will then be fast and accurate. In addition, you will be able to capture in a single stream forms printed using different printers and even those received by fax.
- **Identifiers.** Identifiers are used to detect the form to which a page belongs and to select the necessary template if there are several templates with similar sets of reference elements in the batch. If you process several forms in a single stream, you must put a unique element on each page of a form. This element will tell the program to which form the page belongs. Identifiers can be barcodes, anchors, separators, or static text (for example, a form title or a piece of explanatory text).
- **Graphical images.** You may need to save certain objects as images such as pictures, signatures, seals, or stamps, for example. ABBYY FlexiCapture 8.0 Professional can save image objects and export them to files or databases.
- **Decorative elements.** A form can contain certain decorative elements, for example a company logo.

4.1.2. Marking of data fields

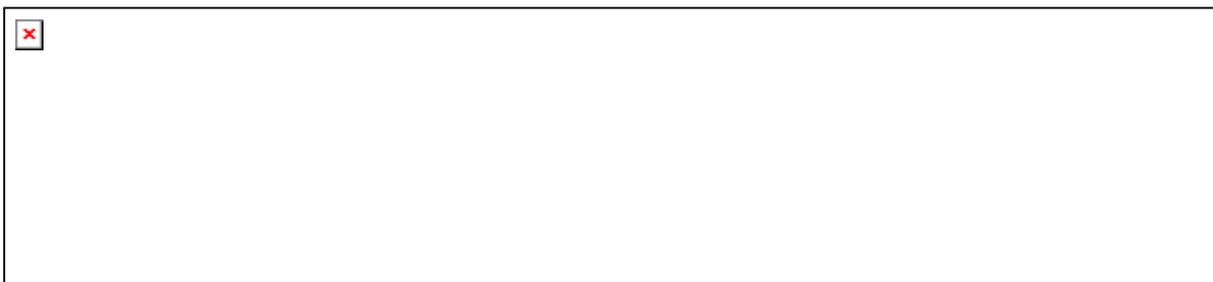
All machine-readable forms can be divided into the following groups by arranging and marking character cells: color drop-out forms, raster forms, and black-and-white linear forms.

- **Color drop-out forms.** Character cells on such forms are white rectangles against a color or grayscale background. Each character cell is intended for one character. You must select the

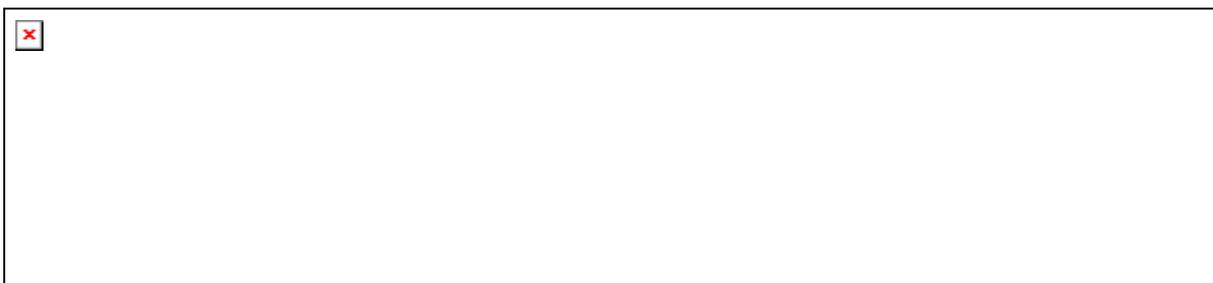
background color and saturation so that the program can easily remove the background during scanning. Ideally, only anchors and filled-out data fields are retained after scanning and despeckling: all other elements must be removed. For this type of processing, you must use a monochrome scanner with a color lamp (red or green), or a color scanner that has a setting to allow the background color to be removed.



- **Raster forms.** To draw character cells, raster forms use dots spaced an equal distance apart. After scanning, these dots are removed from the image using the “Despeckle” option, without any loss of information in filled-out data fields. Character cells can be rectangles with borders made up of raster lines, i.e. sequences of small black dots, or white rectangles on a raster background made up of individual dots.



- **Black-and-white linear forms.** In the case of black-and-white linear forms, field borders are simple lines which remain on the scanned image. During recognition, the program therefore has to first separate the field border from the field content, and then recognize the content. Based on the information about the number of character cells in the data field and the method of separating the cells from each other, the program detects vertical and horizontal lines on the layout and separates them from the field content.



The questionnaire that you have created in ABBYY FlexiCapture 8.0 Professional is a good example of a machine-readable fixed form and is clearly identifiable as a color drop-out form. Please analyze the arrangement of the main elements on this form.

4.2. Creating a project

A **project** contains all of the necessary document capture settings: document templates, image import profiles, program settings, and processed documents.

Documents are grouped into **batches**. The number of batches depends on your processing approach: you can process all documents in one batch, or you can sort documents into batches according to their date of import or scanning date.

Documents are processed in **work batches**. Only work batches are accessible in operator's mode. **Test batches** are used by the administrator to test and adjust templates. The main difference between the two types of batches is that local templates are used to process documents from test batches while published templates are used to process documents from work batches.

A **document** contains images of one or several pages (single-page and multi-page documents) and data extracted from these pages.

A project can contain several templates, in which case documents of different types will be processed within a single project. As a result, you will not have to sort the document to be processed because documents of different types can be captured in a single stream. If, however, document streams do not intersect, you can create a separate project for them.

To begin with, the administrator must create a project and at least one document template.

To create a project, click **Create New...** in the **Open Project** dialog box which appears when you start the program. Alternatively, select **File > New Project...** in the program main window. Specify the folder to which the project is to be stored and enter the project name.

To add batches to the project, right-click in the program main window containing the list of batches (Figure 2) and from the context menu select **New Batch**. You can load images without creating a batch, in which case the program will create a batch automatically.

To view documents added to the batch, double-click the batch name. To return to the list of batches, select **Project > Work Batches List** or press Ctrl+B.



Figure 2. ABBYY FlexiCapture 8.0 Professional main window

Start ABBYY FlexiCapture 8.0 Professional. To create a project, click *Create New...* in the *Open Project* dialog box, which appears when you start the program. Alternatively, select *File > New Project...* in the program main window. Specify the folder to which the project is to be saved and enter the project name. When you click *Create*, the new project will open.

4.3. *Creating a document template*

The most important step in setting up a project is the creation of a template. The quality of data received after forms have been processed depends on the correctness of the template. To create a template, you must specify the:

- **Static elements** on the image: anchors, separators, static text, and barcodes. Select which of these elements are to be used for template matching and document identification. Anchors are detected and marked automatically.
- Location of all **fields**. Fields must correspond to the areas of the image from which data is to be extracted.
- Properties of each field: select the data types to be searched for in every field (this significantly improves the recognition quality) specify which fields must be sent to the operator for verification, etc.
- Rules by which field values are to be checked. Such rules help the program detect documents whose values do not conform to certain conditions, for example where a field value does not correspond to the values of the necessary database.
- Method of data export. Data can be exported to a file or database, or in accordance with a script procedure.

Once the template is created, it must be published in order to use it for subsequent document recognition.

To create a new template, from the menu select **Project > Document Templates...** Click **New...** in the dialog box that opens. This will start the Document Template Creation Wizard. In the **Create New Document Template** dialog box you can specify the template's main properties: its name, description, locale, and writing style. Select the text type: **ICR (hand-printed)** if most document fields are filled out by hand, or **OCR (printed)** if the values in most document fields are printed. In the latter case, select the print type from the drop-down list. The text type specified at this stage will be the default text type, but you will be able to change the text type for individual fields.

Next, load or scan the image on the basis of which the template is created. If your document consists of several pages, load the first page. When adding the rest of the pages, please refer to the recommendations of the section [Creating a template for a multi-page document](#). You can scan the image of a blank page or load it from a file. If you are going to process semi-structured documents, you must use a flexible description when creating a template. If this is the case, select the **Load FlexiLayout** option and specify the path to the AFL file containing the flexible description created in ABBYY FlexiLayout Studio.

You can now select the field types that are to be automatically detected. You can specify checkmarks and text entry fields. Searching automatically for text fields with marking and rectangular checkmarks is highly efficient. However, if text fields on your form contain no marking and checkmarks need to be made against a plain white background, we recommend that you mark such fields manually.

If there are anchors on the image, these will be detected and marked automatically.

1. In the program main window, select *Project > Document Templates...* Click *New...* in the *Document Templates* dialog box.
 2. In the dialog box that opens, specify the parameters for the template, its name and description. In the *Language(locale)* field, specify the language in which the form will be filled out. In the *Writing style* field, select the relevant country. This is because the shape of certain characters, for example, digits, may differ between countries. Select the text type: *ICR (hand-printed)*. Click *Next*.
 3. Select the *Scan* option and scan the blank form without any filters so that the background color is retained. This will make it easier for you to mark data fields on the image of the form (later, when you scan page images for data recognition, you will select a scanning mode which removes the background). The page template will be created on the basis of this image. You can also select the *Load from file(s)* option and load the image from a file. (You can find an image of a questionnaire in <disk name>:\Documents and Settings\All Users\Application Data\ABBYY\FlexiCapture\8.0\Samples\FormDesigner\English\Questionnaire, for Microsoft Windows Vista <disk name>:\Users\Public\ABBYY\FlexiCapture8.0\Samples\FormDesigner\English\Questionnaire). Click *Next*.
 4. Select the *Text* and *Checkmarks* option for automatic marking of text fields and checkmarks. Click *Finish*.
 5. Once the Wizard has completed its work, the *Document Template Editor* window will open. The questionnaire will be displayed in the window, and all text fields and checkmarks will be marked. The names of automatically detected template fields will be displayed in the *Document Structure* window.
- If you switch to Static elements mode (the  button), you will see the layout of static elements. Anchors will also be marked automatically.

4.3.1. Document Template Editor

All the main actions for creating and editing a template are performed in the **Document Template Editor** window (Figure 3), which opens once the Template Creating Wizard, has finished. To open the Document Template Editor from the program main window, select **Project > Document Templates...**, then select the name of the necessary template and click **Edit...**

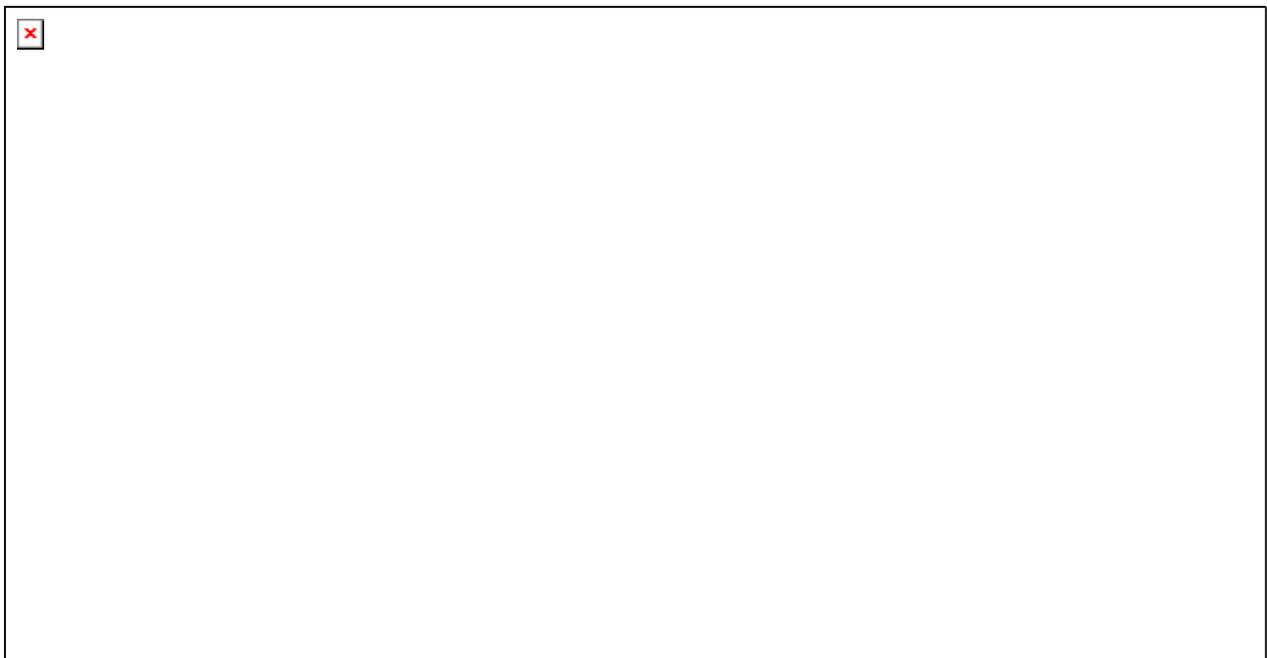


Figure 3. The Document Template Editor

4.3.2. Using elements to mark objects on the form

Once the Template Creating Wizard has finished, the loaded image will be displayed in the Template Editor window. Anchors and data fields of the types you selected during the previous step of template creation will already be marked.

You can automatically mark objects later on by selecting the  tool and clicking on the area of the element to be marked. The program will automatically detect the type and location of the element.

The Document Template Editor provides intuitive and convenient tools to mark fields and static elements. The Editor has two modes:

- Field marking mode (the  tool), and
- Static elements mode (the  tool).

To create a static element or data field manually, switch to Field marking or Static elements mode and click the corresponding button on the toolbar. To create a corresponding element, drag the cursor to draw a rectangle around the necessary object on the form. Alternatively, select the necessary tool and Shift-Click near the object. The area of the data field or static element will now be detected automatically.

Next, follow the list of tools for creating elements of different types:

Fields:



- text entry field



- checkmark



- checkmark group



- barcode



- graphical element (picture)



- table



- group field

Static elements:



- anchor



- separator



- static text



- barcode

A barcode can be both a recognition field and a static element. You must be careful when selecting the marking mode: if data is to be extracted from the barcode, use Field-marking mode. If the barcode is used for document identification and template matching, mark it in Static elements mode.

Created fields appear in the list on the **Fields** tab of the **Document Structure** window. By default, they are assigned names corresponding to the explanatory text. You can rename a field by selecting it in the document structure and pressing F2. To give the field a name corresponding to the explanatory text, select the field, right-click on it, and select **Get Name From Image...** from the context menu. After this, drag your cursor to draw a rectangle over the explanatory text on the image.

The type of field is marked by an icon in the list of fields and by the color of the frame on the image. Static elements are not included in the list.

You can copy elements (even to other document sections), delete, move elements, or change their sizes. If you copy fields, numbers are automatically added to the names of the fields.

To select several elements simultaneously, use Ctrl-Click. The action performed will then be applied to all selected elements. To select elements, use the  tool.

1. Text entry fields are already automatically marked on the form. If you did not select the *Text* option during the final stage of template creation, or if you wish to mark text fields manually, select the  tool followed by the fields in which you want to enter text. A field must include all character cells and some additional space on all sides. Include fields for the fill-in date, name, occupation, country, e-mail address, company name, number of pages processed daily, and the 4 fields for additional information.
2. Please mark the “Your comments and wishes” field. This field has not been detected automatically because it is not marked.
3. If the checkmarks have not been marked automatically, select the  tool and mark all those checkmark boxes that do not belong to any groups. Do not forget to include a little extra space on all sides of the checkboxes. The checkmarks that belong to a group need not be marked individually.
Select the  tool to draw a rectangle around all checkmarks that must belong to the group. Each individual checkmark will be automatically marked and assigned a name, following which all these checkmarks will be united in the group.
4. Switch to Static elements mode. Anchors are already marked on the form. Select the  tool and mark the barcode on the image.

4.3.2.1. Field groups

Fields can be grouped to optimize the document structure and to create repetitive field groups. For example, the city, street, and building number can be combined to form the “Home Address”. If you wish to create fields for the work address, you can simply copy the “Home Address” field group.

To combine fields to form a group, use the  tool.

If a document has repetitive field groups, you can create only one group and then create several *instances* of this group. All field properties and all rules specified within the original group will be inherited by all instances of the group. For more information, please see the [Fields with multiple instances](#) section.

You can also copy a field group. In this case, however, the new group will be independent.

4.3.2.2. Fields without marking

Some fields can have no corresponding regions on the image. The names of unmarked fields are marked with a red asterisk. Such fields can be used to store interim results of calculations, which use values from recognition fields.

Unmarked fields possess all the properties typical of their type: they can be sent to the operator for verification, their format can be checked, and the values of such fields can be exported.

There are two methods of creating a field without marking:

1. In the Template Editor window, select **Template > Create Field** from the menu and create a field of the necessary type. The name of the field will appear in the list and will be marked with an asterisk. A document structure field is created, but not its region on the image.

2. The second method is by removing the marking of a standard field. Select the necessary fields on the image or in the list and from the context menu select **Delete Region**. The marking will be removed and the name of the field will be marked with a red asterisk.

To create a region on the image for a field without marking, select the  tool from the toolbar and frame the necessary region. If there are any fields marked with an asterisk in the list, the program will prompt you to select a name for one of these fields (for example, if you create the list of document fields first and only subsequently specify their location).

4.3.2.3. Marking of Tables

ABBYY FlexiCapture 8.0 Professional allows you to work with tables. Creating fields of the type “Table” makes it much easier to set up, extract, and export data from tables. A Table field is a set of columns of a single type which contains repetitive rows.

The program has special tools for marking tables in a fixed template. These tools will help you draw a table and mark its columns and rows.

Select the  tool to draw the region of the table. Please note that this region must not include the header of the table. Once the table has been created on the image, table-marking tools will appear in the

toolbar. To add separators, mark the table cells using the  tool. To create vertical separators, place the pointer inside the table region, drag the dotted separator to the desired location, and click once. Horizontal separators are created in the same manner while holding down the ALT key. You can also detect separators automatically. To detect separators, select the table you have created and then select either **Autodetect Vertical Separators** or **Autodetect Horizontal Separators** from the context menu.

You can delete any separator using the  tool. Once you have added the number of separators required, mark the columns of the table. To do so, select the  tool and use your cursor to specify the region of a column. Each column contains cells of a single type: text, checkmarks, graphical elements or barcodes. The program will prompt you to select the type of column during marking.

Once the table has been marked, you must specify the recognition properties, verification properties, and data type for each column. Column properties are specified just like the properties of standard document fields.

4.3.2.4. Fields with multiple regions

If your form has data fields whose region consists of multiple parts (for example, tables which start on one page and end on the next page), you can create fields with multiple regions to process such objects on the form.

Values from all regions of a field are joined and exported together as one field. Line folding is used as a separator.

To create a field with multiple regions, create the region of one of the fields, select this region and then select **Continue Region...** from the context menu. Move the created region to the desired location on the page. Select the area on the page where the created region should continue and repeat the procedure as many times as may be necessary.

4.3.2.5. Fields with multiple instances

Your documents may contain repetitive objects – fields or field groups that occur several times in a document and describe similar information, for example similar detail about employees, children, or invoices. To process such objects, you can create fields with multiple instances.

Any field can have multiple instances, and these instances can be spaced any distance apart, even on different pages. Field instances possess identical properties. Fields with multiple instances are exported to separate files or database tables.

Fields with multiple instances are convenient when you have to create repetitive field groups. You can create one field group and then simply create the necessary number of instances.

To create a field with multiple instances, create a single field, right-click on the region of the field, and select **New Instance...** from the context menu. Create as many instances of the field as required and move them to the desired locations on the page.

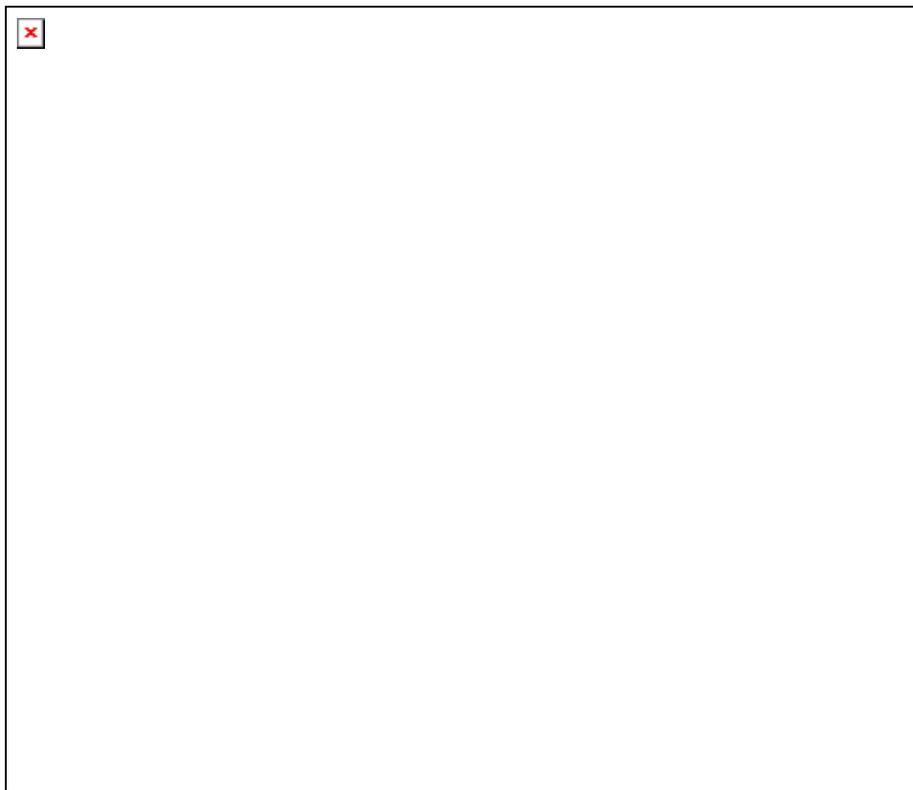


Figure 4. A sample of using a field with multiple instances

4.3.2.6. Excluding a region from recognition

Sometimes you may need to exclude a certain region on a form from recognition, for example, when explanatory text hinders extracting data from a field (see Figure 5). To exclude a region from

recognition, select the  tool and specify the region to be excluded from recognition.

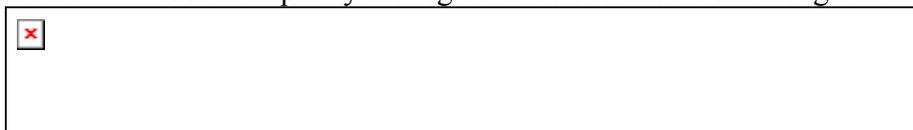


Figure 5. A sample of excluding a region from recognition

4.3.2.7. Deleting fields

To delete a field, select this field and press the **Delete** key or select **Delete** from the context menu. If you want to delete only the region of the field but to retain the field in the document structure, press **Shift+Delete** or select **Delete Region** from the context menu for the field.

4.3.3. Static elements

Static elements mark objects from which data are not extracted. Such elements are used for template matching and document identification only.

To work with static elements, you must switch to Static elements mode by clicking the  button. The marking of static elements is displayed only in this mode.

All types of static element can be used for template matching. However the best results are obtained when a document has special elements – standard anchors: click squares, crosses, or corners. These must be marked automatically or manually as static elements of the *Anchor* type. The shape of anchors can be selected on the **General** tab of the element properties.

You can use static elements of any type as identifiers. The program uses the location of identifiers or their values to determine to what document the current page belongs. If a barcode is used as an identifier, you can specify the values for the barcode. This will help you identify the page quickly and accurately.

To use a static element for template matching and/or identification, you must select the corresponding option on the **General** tab of the Properties dialog box. To open the Properties dialog box of any element, select **Properties...** from the context menu of the element.

Anchor: To use the anchor for template matching (recommended), select the **Use for template matching** option. To use the anchor for identification, select the **Use for template identification** option. Select the type of anchor (a black square, cross, or corner).

Static text: To use the text for template matching, select the **Use for template matching** option. To use the text for identification, select the **Use for template identification** option. If a static text is used for identification, you may specify the value of the text. Please note that you must specify the value of the static text only if it is impossible to identify the page by the location of the text and thus the value of the static text is required (for example, pages differ only in their titles while the location and size of the titles are identical).

Separator: To use the separator for template matching, select the **Use for template matching** option. To use the separator for identification, select the **Use for template identification** option.

Barcode: To use the barcode for template matching, select the **Use for template matching** option. To use the barcode for identification, select the **Use for template identification** option. If a barcode is used for identification, you may specify the value of the barcode. To do so, on the **Recognition** tab specify the barcode type, orientation, and image processing options.

Five anchors have been marked on your form. Please ensure that you have selected the *Black square* type for all the anchors and that the *Use for template matching* and *Use for template identification* options are selected.

For the barcode, select the *Use for template matching* and *Use for template identification* options. If the value of the barcode does not automatically appear in the *Field Value* fields, click *Hint*. The value of the barcode will be inserted.

The barcode on this page corresponds to the EAN 13 type, which must be specified in the *Code type* field on the *Recognition* tab. The *orientation* is *Left to Right*.

Five anchors and one barcode are sufficient to unambiguously identify and match the template, unless you plan to process other documents with an identical arrangement of anchors in the same stream.

4.3.3.1. Peculiarities of barcodes

If a barcode is used as an identifier, then it is an anchor barcode and therefore a static element. Such barcodes must be created in Static elements mode. The Properties dialog box for such a barcode contains two tabs: **General** and **Recognition**

If you plan to extract data from a barcode, such a barcode is a data field and must be created in Field marking mode. The Properties dialog box of such a barcode contains all tabs typical of data fields, namely **General**, **Data Type**, **Recognition**, **Verification**, and **Rules**. The value of such a barcode will be recognized and, depending on the settings, sent to verification.

4.3.4. Field properties

It is very important that field properties are specified correctly. Field properties affect the quality of recognition and determine if the values of the field are to be exported and sent to the verification operator. Certain properties are particularly important for data recognition. A good example is the type of marking of a text field. You must specify this property correctly so that the marking that is not removed during scanning is separated from the characters. If you do so, the recognition result will contain only the text which was entered in the field, and no excess objects will be captured.

If field properties are specified correctly, the results of data recognition will be very good and no extra work for operators will be created. When you specify field properties, you can keep document verification to a minimum because values will be automatically corrected.

When you have created a field of a certain type, default properties are assigned to the field. To change the properties of an element in the dialog box, select **Properties** from the context menu of the element. Every type of field has its specific list of properties.

4.3.4.1. Common field properties

The **Properties** dialog box of a field of any given type contains the **General** tab (Figure 6). On the **General** tab, you can specify the name of the field (**Name**) and provide a description. When you create a field, the program will generate a name for it automatically, based on the explanatory text closest to the field. You can change this default name to any other name that is convenient. **Caption** is the name of the field as it will be displayed in the data form. The field type is specified on this tab using the icon to the right of the field name.

In addition, the following options can be selected on the **General** tab:

- **Export field value.** Clear this option if you do not need to export the value of this field. You may need to clear this option, if, for example, the value of the field is used to get the value of a calculated field (see the [Rule-based checks](#) section) and only the final result has to be exported.
- **Read only.** Select this option to make the field read-only. The operator will not be able to change its value. You can select this option for fields whose value must be calculated automatically, with the help of a rule. For example, you can do so for a field where the sum or merged values of other fields will be stored (see the [Rule-based checks](#) section).
- **Show in data form.** Uncheck this option if do not want this field to appear in the Data Form.
- **Cannot be blank.** Check this option if the field must be filled out.

- **Index field.** Select this option if you plan to use this field for document indexing. If you do so, the value of the field in each document in the list will be indexed, and the operator will be able to use the value of this field for sorting and searching documents.



**Figure 6. The General tab of the Properties dialog box
(for a text entry field)**

When you create fields, they are automatically assigned names corresponding to the explanatory text. Please ensure that you name the fields you created correctly. If necessary, you can change the name of a field on the General tab of the Properties dialog box.

If you wish, you can provide a description for the element.

The *Export field value* and *Show in data form* option are selected for all fields by default. Please make sure that the *Export field value* option is selected for all text fields, checkmarks that do not belong to any groups, and checkmark groups. Note that properties of checkmarks belonging to a group are identical for all checkmarks in the group. Therefore these properties are specified once for the entire group.

4.3.4.2. Data types

A **Data type** determines possible values of a field and its allowed format. If a value entered in a field does not correspond to the specified data type, the operator will obtain an error message when verifying the field. A data type of a text usually describes the aggregate of allowed values, for example, date, time, address, amount of money, etc. A data type of a checkmark represents the values acquired by the field if the checkmark is checked or not checked.

4.3.4.2.1. *Data types of a text entry field*

It is very important that the *data type* for text fields is specified correctly. Specifying the data type tells the program what kind of data is expected in the field: digits, or letters of a certain alphabet, or characters from a certain set, a date, etc. The program has a flexible mechanism for specifying data types. The user is provided with a ready-made set of data types that includes the most common types. In addition, users can create their own data types to perform a specific task.

When specifying data types, you can set up the value format check and value constraints, for example, maximum and minimum values for a number or a time interval for a date.

The data type is specified on the **Data Type** tab of a field's **Properties** dialog box (Figure 7).

Select the desired category from the **Content** list. In the **Details** field, you will see a description of one of the data types of this category (either the default one or the one that you set up earlier).

If **Process value as text** is selected, the values of fields with any type of content will be processed and exported as text.

If you wish to change the recognition language or set a narrower data type, click the **Edit...** button to the right of the description.

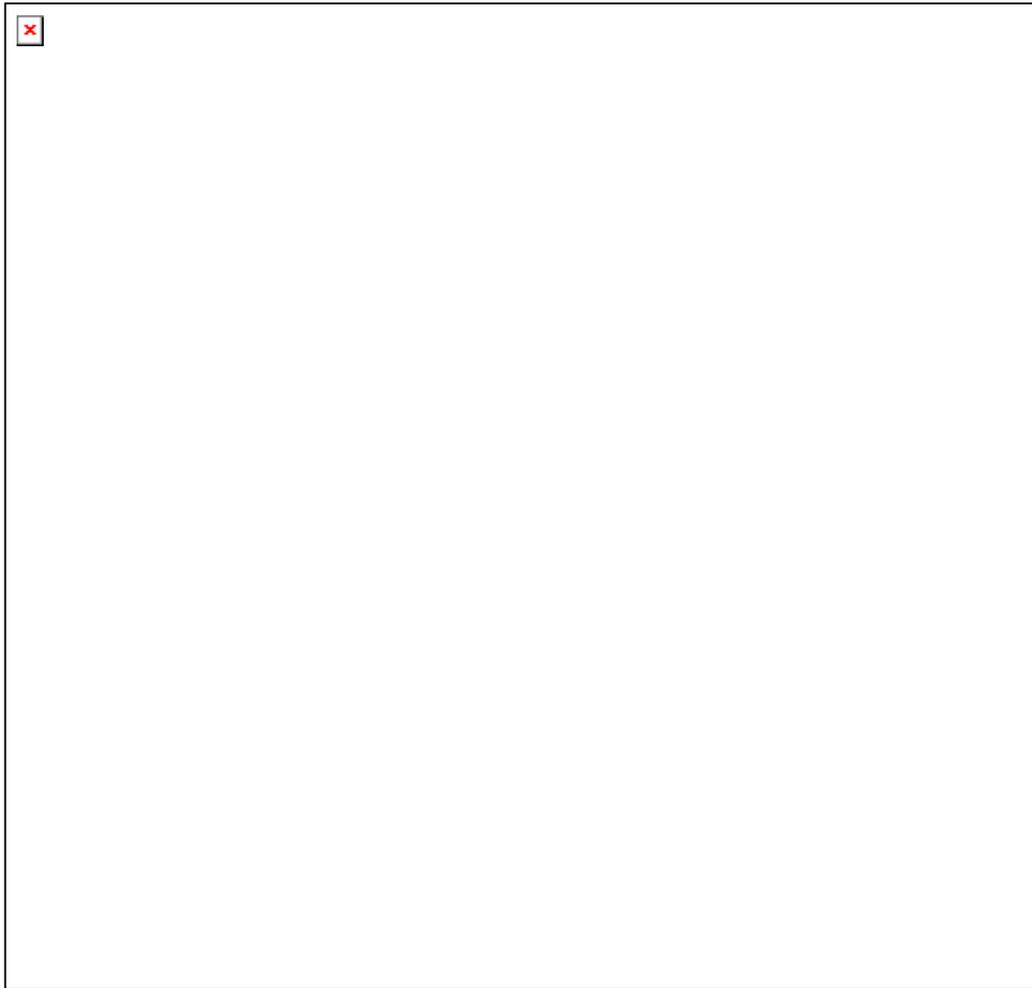
The following options are available for the standard data types (**General** is selected in the **Content settings** list):

- For Text, you can select multiple recognition languages (the "...” button), use the built-in and/or custom dictionary (the program will check field values against the dictionary you specify)
- For Number, you can select **Integer**, if you expect the value of the field to be an integer
- For Date, you can select the order of date components, specify if the month can be written as a word, and specify if the date may include time or day of the week
- For Address, Name, and Code, you can specify a custom dictionary

The special data types (**Special** is selected in the **Content settings** list) contain predefined data types from which you can select the data type most suitable for a given field. Consult the descriptions displayed at the bottom of the dialog box when selecting a data type.

If none of the data types in the list suits your purposes, you can create your own data type:

1. On the **Data Type** tab of the field properties dialog box, select one of the items from the **Content** list. You can select any item that best corresponds to the nature of your data type. This choice will not in any way affect the creation of the data type: the newly created data type will only be stored under this category.
2. Click the **Edit...** button to the right of the **Details** field. In the dialog box that appears, select **Special** in the **Content settings** field. Click **New...** to the right of the **Select data type** field.
3. Follow the prompts of the data type creation wizard.



**Figure 7. The Data Type tab of the Properties dialog box
(for a text entry field)**

For any data type, the program can automatically process the entered values: remove excess spaces, change letters to uppercase or lowercase, or automatically replace specified characters or text fragments. To enable automatic processing of values, click the **Edit...** button next to the **AutoCorrect options** field. In the dialog box that opens, select the required automatic processing options (Figure 8).



Figure 8. Auto replace dialog box

You can also set up a checking procedure for recognized values to check if they belong to a certain interval. To specify an interval, click **Edit ...** next to the **Validation** field (Figure 9).

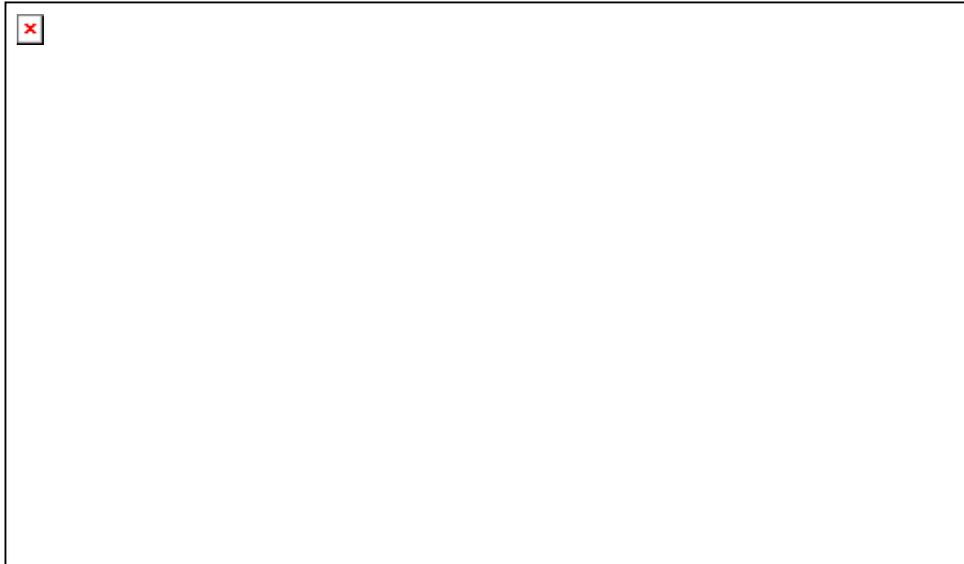


Figure 9. Value checks dialog box

Specify a data type for each text field. For the *First and last name* field, you must select the *Name* type and specify the correct language. For the *Processing volume* field where the number of pages is indicated, select the *Number* type (the format is an integer).

For the "Fill in date" field, select *Date* from the *Content* list. Then select from the list or describe the date format. If you wish to check that the date is within the specified time interval, click *Edit* in the *Validation* area and select the *Date must be within the following interval* option in the dialog box that opens. You may specify a fixed or relative time interval. For example, you can specify the following time interval: 90 days before the current date and 0 days after it. In this case, no error message will be displayed for questionnaires which were filled out no earlier than 90 days before the current date and no later than the current date.

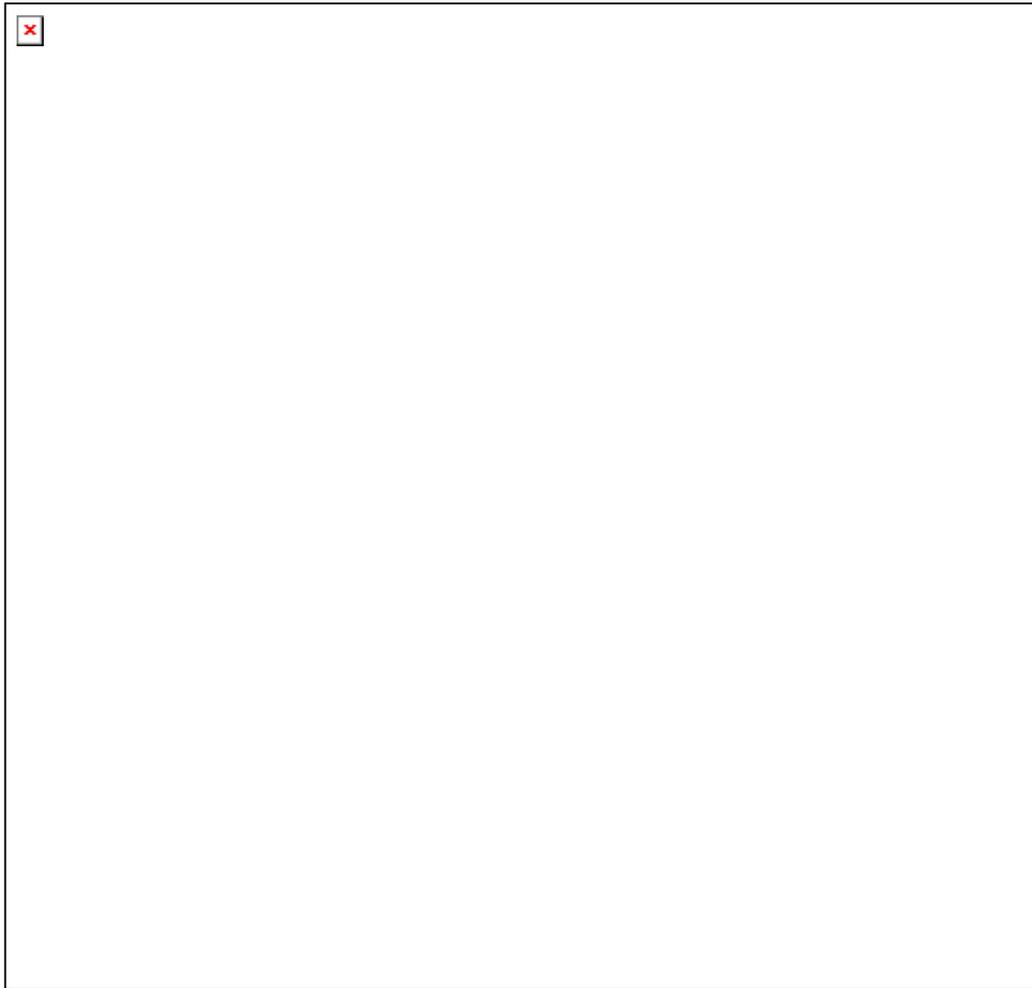
4.3.4.2.2. Data types for checkmarks

When selecting a data type for checkmarks on the **Data Type** tab, you specify the values to be assigned to fields if their checkmarks are selected or not (Figure 10).

You can assign the following values to checkmark fields:

- *Yes* if the checkmark is selected, *No* if the checkmark is not selected;
- *1* if the checkmark is selected, *0* if the checkmark is not selected;
- *Checkmark name* if the checkmark is selected, *Empty* if the checkmark is not selected;
- You can also select the *Custom* option and provide your own values.

Note. If checkmarks are combined into a group, they possess common properties. Properties are specified only for the entire group, but checkmark values are also specified in the group properties.



**Figure 10. The Data Type tab of the Properties dialog box
(for a checkmark which does not belong to a group)**

4.3.4.2.3. Data types for a checkmark group

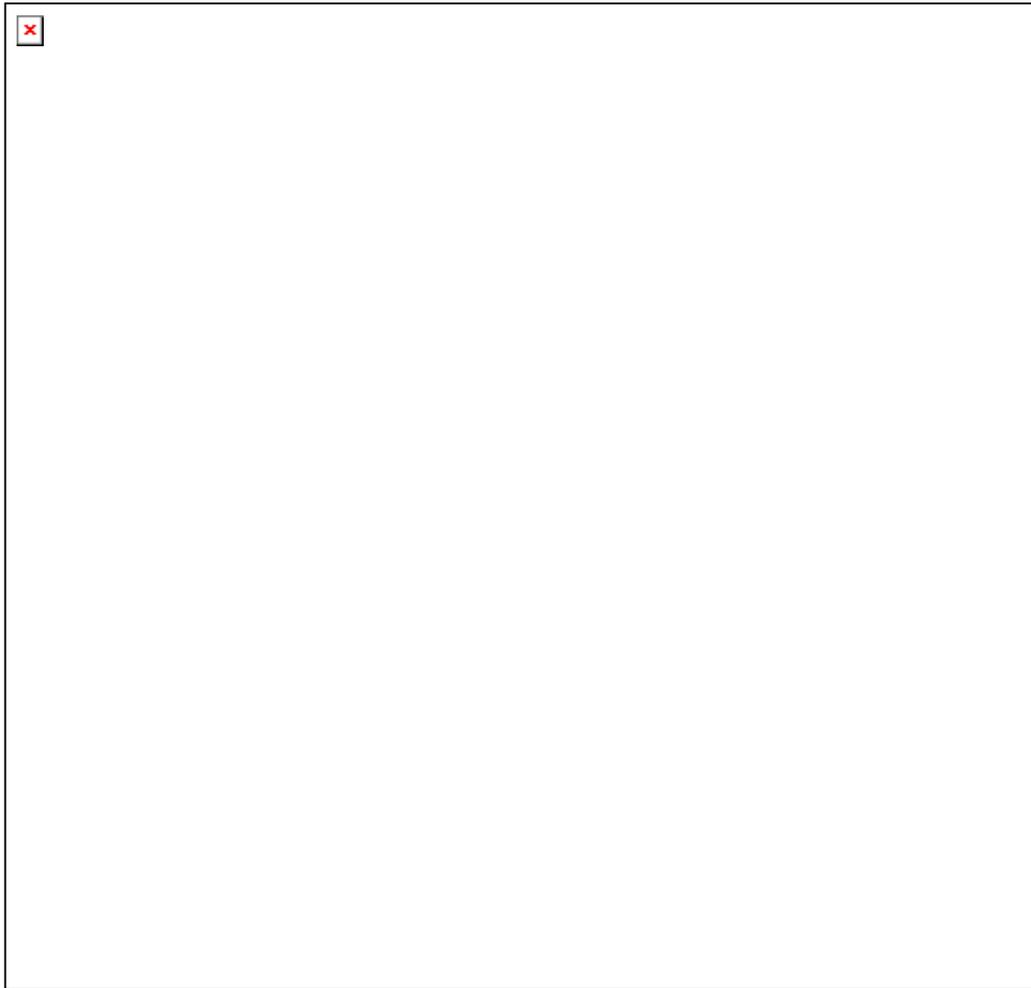
On the **Data Type** tab of the **Properties** dialog box for a checkmark group, you can see a list of all checkmarks included in this group (Figure 11).

Select **Allow empty selection**, if at least one checkmark field in the group must be checked.

If multiple checkmarks in the group can be selected, you must select the **Allow multiple selection** option.

You can specify what values must be exported if no checkmark field has been selected or if more than one checkmark fields have been selected. Select either **<Empty selection>** or **<Multiple selection>** from the list and click **Edit...** IN the dialog box that opens, type the desired value in the **Exported value** field. If no values are specified, the program will export an empty string whenever no checkmark fields are selected and the values of the selected checkmark fields separated by commas, if more than one is selected.

If **Treat validation error as warning** is selected, the program will display a warning message instead of an error message for validation errors.



**Figure 11. The Data Type tab of the Properties dialog box
(for a checkmark group)**

Specify properties for checkmarks and checkmark groups. Select the <Checkmark name>/Empty method of checkmark values conversion.

For the "Types of documents to be processed" checkmark group, select the *Allow empty selection* and *Allow multiple selection* options. Specify the value to be exported if no checkmark field is selected, for example, "none selected". Specify the value to be exported if multiple checkmark fields are selected, for example, "more than one selected".

Do the same for the other checkmark groups.

4.3.4.3. Field recognition properties

ABBYY FlexiCapture 8.0 Professional allows you to specify recognition properties for each individual field. If you specify these properties correctly on the **Recognition** tab of the **Properties** dialog box, the recognition quality will be much higher and errors will be much less likely. Recognition properties of different field types differ. Below there follows a detailed description of these properties for all field types.

4.3.4.3.1. Recognition properties of a text entry field

Select the **Don't recognize (Key From Image field - will be entered manually)** option only if you do not want to recognize this field and its value is to be manually entered by the operator (for example, if it is too difficult to recognize the field because the letters are italicized). If this is the case, you do not need

to specify any other recognition properties because the field will not be recognized and the verification operator will be prompted to enter the value of this field manually.

Select the text type: **ICR (handprinted)** or **OCR (printed)**. For printed text, select the print type (typographic, matrix printer, typewriter, etc.). To specify multiple text types or use a template, select the **Advanced** option and click **Modify...**

Select the marking type using the marking samples from the drop-down list. If the marking is removed during scanning, select the monospaced (Gray boxes) type. If the marking is not removed and there are character cells, specify the total number of character cells (the number of cells can be detected automatically).

You can select a **letter case** for letters, which will make all the recognized letters either upper case or lower case. If a field may contain both upper and lower case letters, keep **Auto**.

Select the orientation of the text (vertical or horizontal).

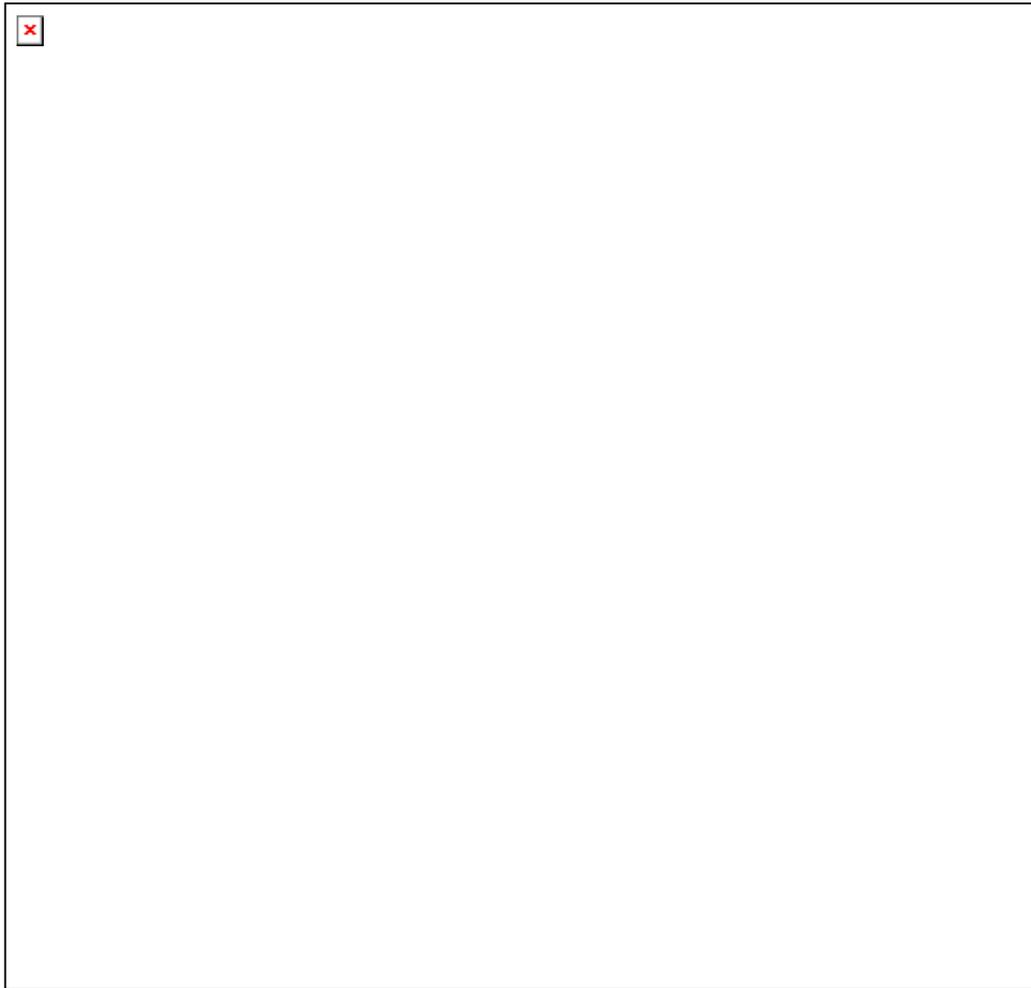
Select **One line** if the field has only one line. If the value of a field is always a single word (i.e. it does not contain any blank spaces), select **One word**.

You can select the letter case, in which case all recognized letters will be changed to upper or lower case.

Specify the image preprocessing options. The program can despeckle the image (you can specify the garbage size), invert the image, and remove the image texture.

Note. If the **Despeckle** option is selected, the garbage from the image will be cleared automatically by default. If you wish to specify the size of the garbage, select the option **Clear the garbage of specified size only** and enter the necessary size.

Select the text orientation: vertical or horizontal.



**Figure 12. The Recognition tab of the Properties dialog box
(for a text entry field)**

1. Specify recognition properties for all text fields on your form. The text type for all fields must be *ICR (hand-printed)*, marking type – *Char box series*, number of cells must be detected automatically, and the text orientation must be *Horizontal*. There are no multi-line fields on the form, so the *One line* option must be selected for all fields. You may select the *One word* option for fields where no blank spaces are allowed, for example, the E-mail field.
2. Specify the letter case. You can retain the *Auto* value of the *Letter case* property for all text fields.
3. The “Your comments and wishes” field is a field for manual entry. The text in this field may be italicized and it may be impossible to recognize it automatically. We therefore recommend that you select the *Don't recognize (Key From Image field - will be entered manually)* option for this field. The operator will then have to enter the value for this field.
4. You may retain the default values for all other fields.

4.3.4.3.2. Recognition properties of checkmarks and checkmark groups

For checkmarks to be recognized correctly, you must specify the type of checkmark recognition known as the **Checkmark type**. If a checkmark is limited by a square, select the **Square** type. If a checkmark is to be placed on a white background without any limiter (or with a limiter which is removed during scanning), select the **Without Frame** type. If the checkmark type is **Auto**, the program will automatically detect the shape of the checkmark. Please note that in this case the checkmark on the form must not be selected, because the program detects whether or not the checkmark is selected by

comparing the image of the checkmark on a processed document against the image of the blank form used to create the template.

You can allow corrections for certain checkmarks. If the person filling out the questionnaire has selected a checkmark by mistake, he or she can just erase this checkmark. Checkmarks that are completely erased will be considered not selected. If, however, you selected the **Auto** type for a checkmark, no corrections are allowed.

You can specify image pre-processing options for checkmark recognition. These options are similar to those of text fields.

If checkmarks are combined into a group, they possess common properties. Recognition properties are specified in the same way, but for the entire checkmark group.



**Figure 13. The Recognition tab of the Properties dialog box
(for a checkmark)**

Specify the properties of checkmarks and checkmark groups.

Select the *Square* checkmark type:

- for checkmarks which do not belong to a group;
- for checkmark groups.

To allow checkmark corrections, select the *Allow corrections* option.

4.3.4.3.3. *Barcode recognition properties*

The properties used to recognize a field barcode are similar to those for a static barcode. For a field barcode, you must specify the barcode type, orientation, and image despeckling options. The only difference is that the operator can enter the value of the field manually. To do so, select the **Don't recognize (Key From Image field - will be entered manually)** option.

4.3.4.3.4. *Picture recognition properties*

On the **Recognition** tab, select the **Exclude from recognition** option for the Picture field if the region of the picture must be excluded from the recognition regions of the other fields. (This option is only available for a Picture fields and ensures downward compatibility with FormReader 6.5 templates.)

4.3.4.4. Verification parameters

Verification is the process of checking recognized data and is performed by the operator. When creating a template, you can set up verification options on the **Verification** tab of the fields properties dialog box (Figure 14). Uncertainly recognized characters are marked by the program and sent to the operator for verification. You can however set up field properties so that the field is sent for verification even if it does not contain any uncertain characters, or, alternatively, that the fields are not sent for verification even if they contain some uncertain characters. Select **All** (obligatory verification) if an error in the value of this field is unacceptable.

The operator sees a recognized character and its image and either accepts the recognition result or corrects it.

If you wish to include all the characters of a field in group verification, select the **Include in group verification** option. To include a field in field verification, select the **Include in field verification** option. Any fields can be included in group verification. However, we recommend that you only send checkmarks and text fields containing nothing but digits and separators (dots, commas, dashes) for group verification. Letters should be checked taking into consideration the context of the field (for more information about group and context verification, see the [Verification](#) section below).

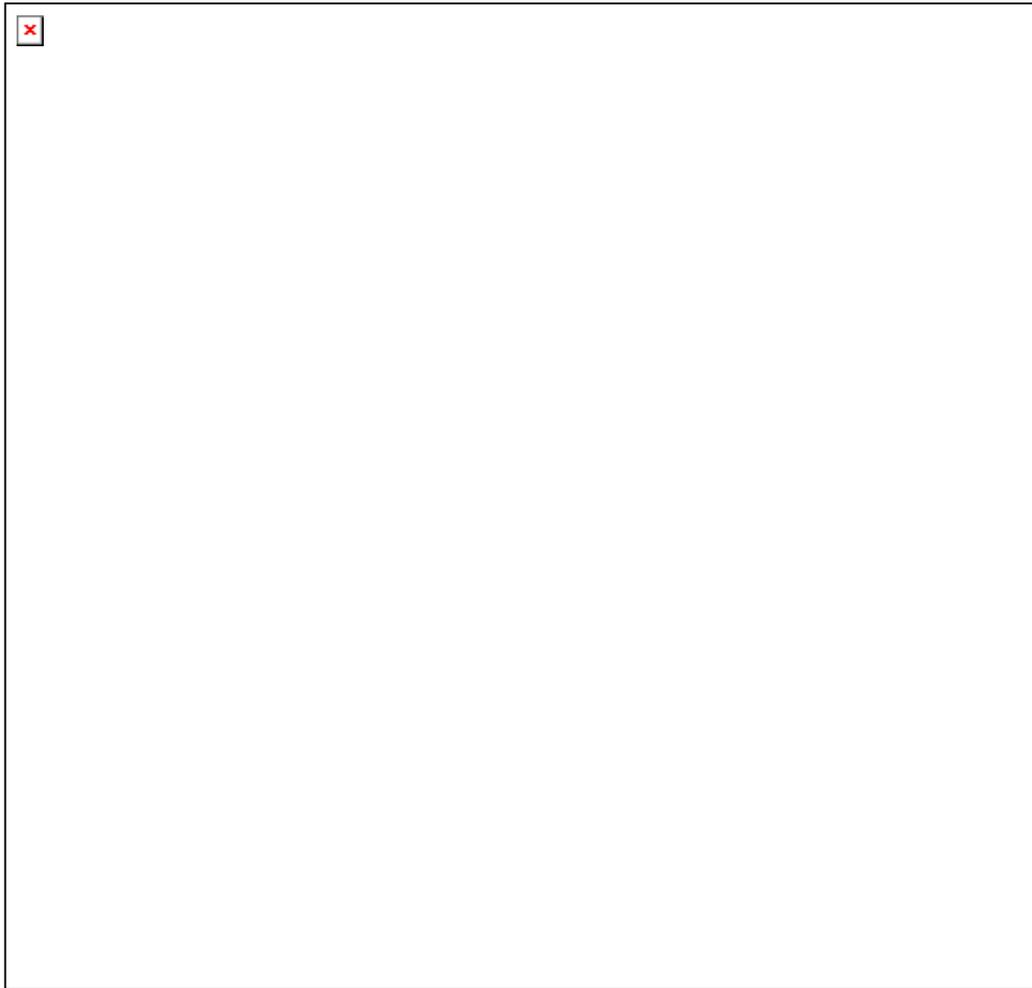


Figure 14. The Verification tab of the Properties dialog box

1. Set up verification options for the data fields. The “Verify *uncertainly recognized* characters” value is selected by default, which means that uncertainly recognized characters in a field will be sent for verification. Keep this value for all fields except, for example, the *First and last name* field and select *All* values for this field. If you do this, all characters of the *First and last name* field will be sent for verification regardless of the certainty of their recognition.
5. Select the *Include in group verification* option for all checkmarks and text fields which must be filled with digits (the *Fill in date* and *Processing volume* fields).
6. Select the *Include in field verification* option for fields that you wish to check by means of context verification (for example, name, occupation, and company name).

4.3.4.5. Picture export options

For a Picture field, you can select export options, including file type, quality, color, and resolution. The following options can be selected on the **Export** tab of the field properties dialog box (select **Properties...** from the field’s context menu):

- **File type** (TIFF, JPEG, BMP, JPEG2000, PCX packbits, PNG).
- **Quality** (best, high, normal, low). The option is available for TIFF, JPEG, and JPEG2000 files.
- **Color type** (color, grayscale, black and white).

To specify a resolution value for a picture, select **Change resolution to** and then select the desired value from the list.

4.3.4.6. Rule-based checks

Rules are used to check recognition results automatically. Rules, like data types, allow you to specify data constraints. i.e. specify the requirements that values of certain fields must meet. If the values in filled-out documents do not meet these requirements, such pages are marked with a flag and a corresponding message. The main purpose of using rules is to check the data integrity in a document. Rules can also be used to process entered data, for example to merge the values of several fields or to substitute recognized values with corresponding values from a database.

ABBYY FlexiCapture 8.0 Professional allows you to create rules of the following types:

- **Check Sum** – checking the sum of values of several fields. This sum is compared against the specified number or a value in another field. For example, if the document contains the salary and the bonus payment of an employee and there is a field for the total income of this employee, you can create a rule that will check if the sum of the salary and the bonus payment is equal to the total income. If it is not, the program will display an error message.
- **Compare Fields** – comparing the values of several fields. This rule may be used if a document contains several fields whose values must be identical. If they are not, the program will display a rule error message for the document.
- **Database Check Rule** – checking entered values against certain values in a database.
- **Merge Fields** – merging the values of several fields. Values of merged fields may be separated by dots, blank spaces, or other separators. For example, it may be convenient not to recognize the data as a single element but to merge the values of separately recognized fields “Day”, “Month”, and “Year”. Dots may be used as separators of the merged value. The result of such merging may be stored in any field of the template. It is particularly convenient to use fields without marking to store such results (see [Fields without marking](#)).
- **Sum in figures - sum in words** (Russian language only). This rule is only available if Russian is selected as the recognition language for the template.
- **Script Rule** – the user can describe the constraints with the help of a script.

Rules are specified on the **Rules** tab of the **Properties** dialog box (Figure 15). A rule can be assigned to one field or to several fields simultaneously.

You can assign the error or warning status to a rule. If the requirement is not satisfied, a field will be marked with a red flag in the case of an error or with a yellow flag in the case of a warning.

If the program tries to export a document containing rule errors, a corresponding warning will be displayed.

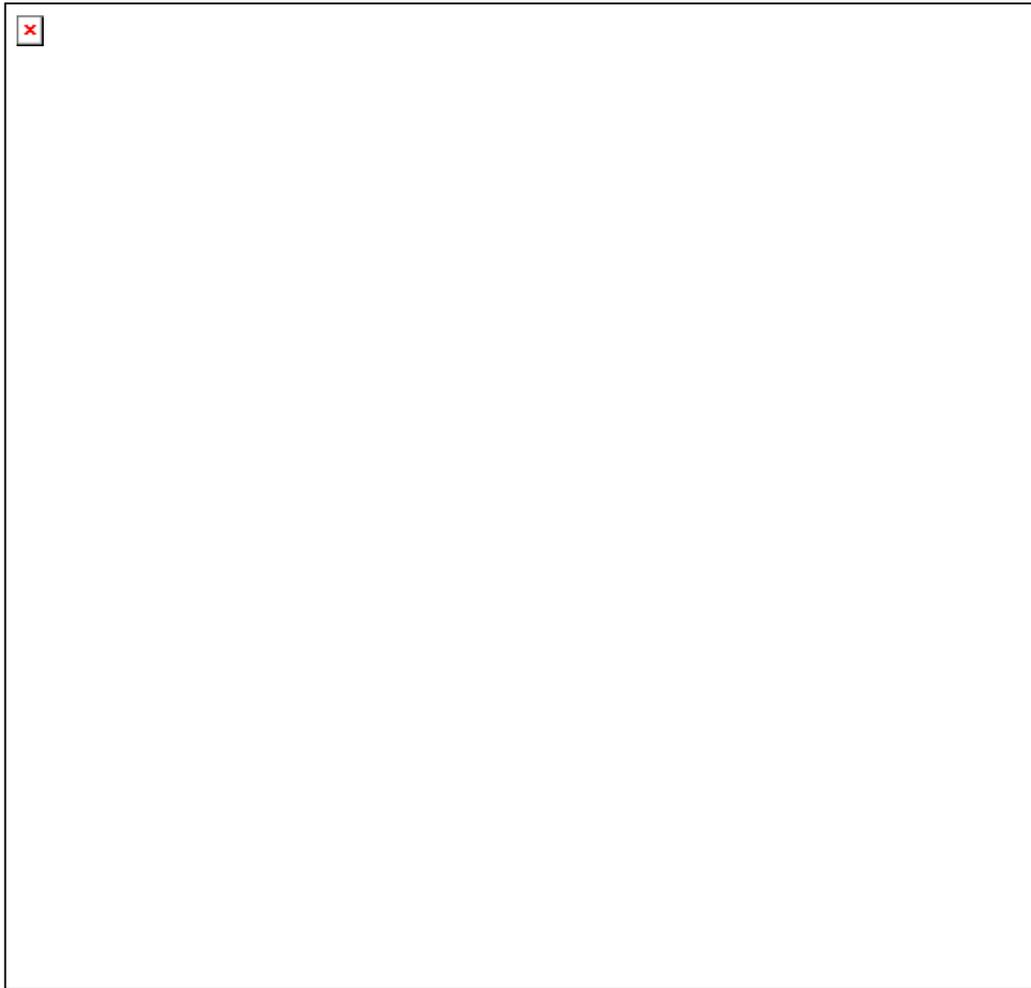


Figure 15. The Rules tab of the Properties dialog box

Let us analyze a sample rule. We will describe the fill-in date of the questionnaire as the merging of the *Day*, *Month*, and *Year* fields. To do this, perform the following:

1. Delete the region of the date field on the form image but do not delete the field itself. To do this, select *Delete Region* on the context menu of the *Fill-in date* field. The marking will be removed from the image, but the field itself will remain in the document and will be marked with a red asterisk.

2. Create the *Day*, *Month*, and *Year* fields.

3. Specify the properties of these fields. On the *General* tab, do not select the *Export field value* option because we will export only the entire date (make sure that the *Export field value* option IS selected for the *Fill in date* field).

4. On the *Data Type* tab, specify the data type for each of the fields. You must select the *Number* data type, and the *integer* format. In addition, specify constraints for each of the fields by clicking *Edit* in the *Validation* area:

- Day: from 1 to 31,
- Month: from 1 to 12,
- Year: depending on the period in which the questionnaire is filled out, for example from 2000 to 2020.

5. On the *Recognition* and *Verification* tabs, specify properties similar to those that we specified for the *Fill-in date* field.

6. Open the *Properties* dialog box of the *Fill-in date* field and go to the *Rules* tab to start creating a rule.

7. Click *New Rule...* Select the *Merge Fields* rule type.

8. Assign a name to the group, for example, *Merge Fields – Fill-in date*. Select *Error* to assign the Error status to the rule. The error will occur if a field participating in the rule has not been detected. Click *Next*.

9. Add the *Day*, *Month*, and *Year* fields to the *Fields* list by clicking *Add*. In the *Result field* field, select the *Fill-in date* field. Use dots as separators.

We set up the check if the date in the *Fill-in date* field lies within the specified time interval. The check will now be performed for the value received by merging the values of three fields.

This completes the step for creating fields and static elements on the form and for specifying their properties. Please make sure that you have created all of the necessary elements and specified their properties correctly. If, after the template matching, you are not satisfied with the recognition results or if some of the elements are located incorrectly, you can always return to template editing and change the marking or properties of the necessary fields.

4.3.5. **Creating a template for a multi-page document**

ABBYY FlexiCapture 8.0 Professional allows you to create templates for multi-page documents. A template can consist of any number of *sections*, and each section may contain one or multiple *pages*. When creating a multi-page document template, you must specify the sequence of sections, their total number, and the rules for assembling pages into documents.

Once the Template Creation Wizard has finished, a simple template containing no sections is created. You can add pages to the template. When you add a page, its image is added to the image area in the Template Editor window, and all fields created on the page will be represented in the total list of fields. In this case a document consists of one section, which may contain multiple pages.

If the document contains fields or tables which continue on the next page, or repetitive blocks, you must add pages to a single section. The same method is used when loading a multi-page flexible description (see [Setting Up the System for Processing Flexible Documents](#)). All fields of a flexible description must belong to a single section of a document template.

An alternative method is to add sections to a template. In the simplest case one section contains a single page. This method can be used if the field sets on pages are independent. This method also makes the document structure easier to understand because you can see to which section fields belong and you can specify a structure for document assembly. For example, your document may contain 3 pages, with the 1st and 3rd pages occurring only once but the 2nd page repeating between 2 and 5 times. Here you need to create a separate section for each page and then specify the document structure, i.e. the sequence and number of reiterations of the sections.

A more complicated example would be a template consisting of several sections where some of these sections contain more than one page. This could be a template consisting of fixed sections and a flexible multi-page section, or a template describing documents containing a double-sided page that repeats several times. Let's assume that we need to create a template for a document containing a title page (page A) and some two-side pages (page sequence BCBC...). We create Section A containing page A, and Section BC containing two pages, B and C. The document then has the following structure: Section A repeats once, and is then followed by Section BC which may repeat, for example, between 3 and 7 times.

If a template consists of multiple sections, you can not only specify the sequence and the number of reiterations of the sections but also set up the *key fields* check. To do this, you must specify key fields in each of the document sections. The values of these key fields must be identical. If the name of the person who fills out the form appears on every page of the document, you can use it as a key field. Alternatively, you may use an identification number.

When processing batches, the program tries to assemble successive pages into documents and checks assembly rules. The values of key fields are checked at this stage. If key fields within a document are not identical, an error message will be displayed. This can sometimes happen if the pages were mixed up

during scanning. If this is the case, you only need to change the order of pages and the requirement of the assembly rule will be met.

To add a page to a section, from the Document Template Editor menu select **Template > Add Page...**, or from the context menu of the image, select **Add Page...** . Next, load the image of the new blank page and select types of objects that must be detected automatically on the page.

To add a new section, from the Template Editor menu select **Template > Add Document Section...** The Section Creation Wizard will open which will help you set up all of the necessary parameters. Follow the Wizard instructions and perform the following actions:

- Specify the section name;
- Load the image of the blank page (scan or load from a file) or load the flexible description file;
- Select the types of objects that must be detected automatically.

The new section and its fields will be displayed in the document structure window.

To view the document structure created during template creation, go to the **Pages** tab of the **Document Structure** window located in the right-hand section of the Document Template Editor window. The document structure is represented on the **Pages** tab as thumbnails. On this tab you can change the number of reiterations of sections in the document (to do this, enter a new value, number or interval in the box to the right of the section name), attach appendix pages, change the sequence of pages in a section, and move pages from one section to another using drag-and-drop.

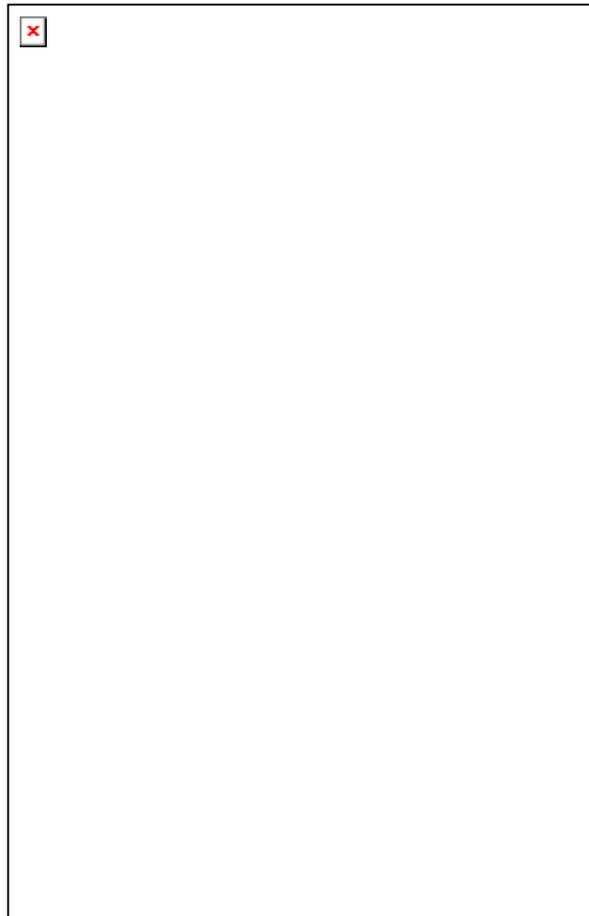


Figure 16. Document structure, page thumbnails mode

You can also specify the sequence and number of reiteration of the sections by selecting **Template > Document Template Properties...** in the Template Editor window.

On the **Assembly** tab (Figure 17), specify the minimum and maximum number of reiteration of the sections in the document (1 by default).

If you wish to check values of key fields, select the **Check equality of key fields** and specify the key field on each of the pages.

In some cases, you may need to disable checks for the order of sections in a document (e.g. the order of sections may be irrelevant for document assembly). Select **Disable section order check** if no section order check is needed. The program will check to make sure that all the sections are present, but it will not check their order.

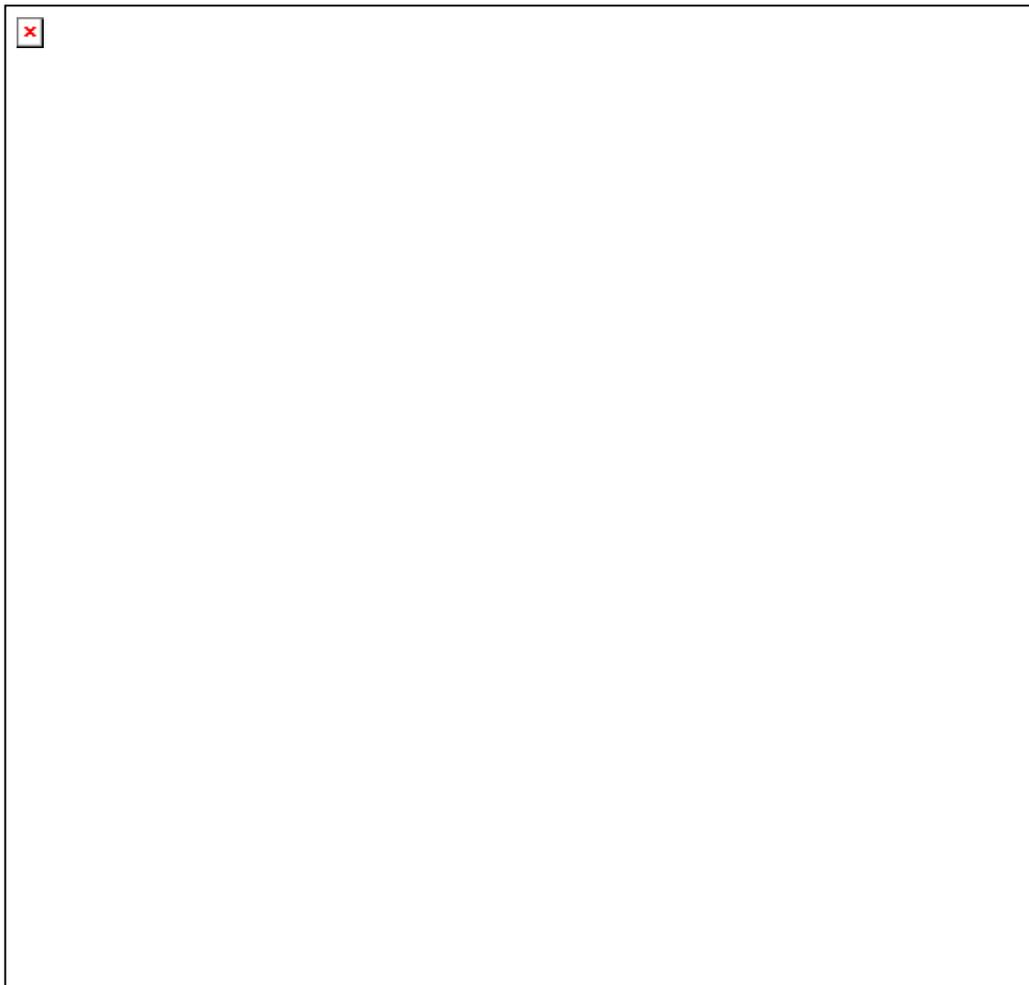


Figure 17. The Assembly tab of the Properties dialog box

Our form is not a multi-page document. The template will consist of only one page and no assembly rules are required. On the *Assembly* tab of the *Document Template Properties* dialog box, you may specify that only one section that repeats only once is displayed.

4.3.6. Creating a template for a document with annex pages

ABBYY FlexiCapture 8.0 Professional allows the creation of templates for documents with annex pages.

Annex pages are additional pages that may be included in any document. They do not contain any recognition fields and you do not have to match a template to them. However, they are taken into account when assembling documents. For example, an application for credit is a fixed form. A certificate from the workplace written in a free-form style is attached to the application. This certificate may be processed as an annex page.

To create a template for a document with annexes:

- Go to the **Pages** tab in the **Document Structure** window (the right part of the Template Editor) and select the **Enable annex pages in document** option. Enter the numbers or interval of annex pages in the box that appears on the right of the section name. (Figure 16)
- Alternatively, open the document template properties dialog box (**Template > Document Template Properties...** in the Template Editor window) and go to the **Assembly** tab. Select the **Enable annex pages** option. Specify the minimum and maximum number of annex pages (Figure 17).

To save the image of an annex page, you must specify the corresponding image saving options when setting up the export (on the **Images** tab of the **Export Settings** dialog box). An image can be saved in a graphical format or searchable PDF.

4.3.7. Setting up data export

To set up the method of saving the data obtained when processing paper documents, the administrator must set up data export for each document template. Four types of export are available: export to a file of a specified format, export to a database, export to a Microsoft SharePoint library, and custom export that uses a script. To set up data export, select **Template > Export Settings...** in the Document Template Editor window and specify the necessary options in the **Export Settings** dialog box.

Export type (file, database, SharePoint, or custom) is specified in the **Export type** field. The value that you specify in this field determines the subsequent settings.

4.3.7.1. Export to files

To set up export to file, select **Export to files** in the **Export type** field (Figure 18).

On the **Destination** tab, select the folder in which export files are to be stored. You can export documents from one batch to a single file (select **Create separate folder for each batch**) or each document to a separate file (select **Create separate folder for each batch document**). If you do not select any of these options, all documents will be exported to a single file.

Select the **Overwrite existing file** option if you wish to allow overwriting of the export file.

Specify how the names of export files should be generated: click the **File Naming Options...** button and select the desired naming options in the dialog box that appears.

On the **Format** tab, select the file type (the following formats are supported: DBF, TXT, XLS, XML) and specify additional export properties for the selected file type. You can also specify the text encoding.

The **Images** tab contains image export parameters. To save images, select the **Save document images** option. Select the folder and file name to which the processed images will be saved. Alternatively, select the **to data folder** option to save the images in the same folder as where the data is saved.

Select the format in which images are to be stored. If you select PDF and select the **Create searchable PDF** option, the program will recognize the entire text of the document and save the recognized text in

the selected format. In this case you can specify a recognition language: either keep the language specified in the template or select one or more languages from the list (**Select** button).

If you wish to change the resolution of initial images, for example to reduce the size of stored data, select the **Change resolution to** option and enter a new resolution.

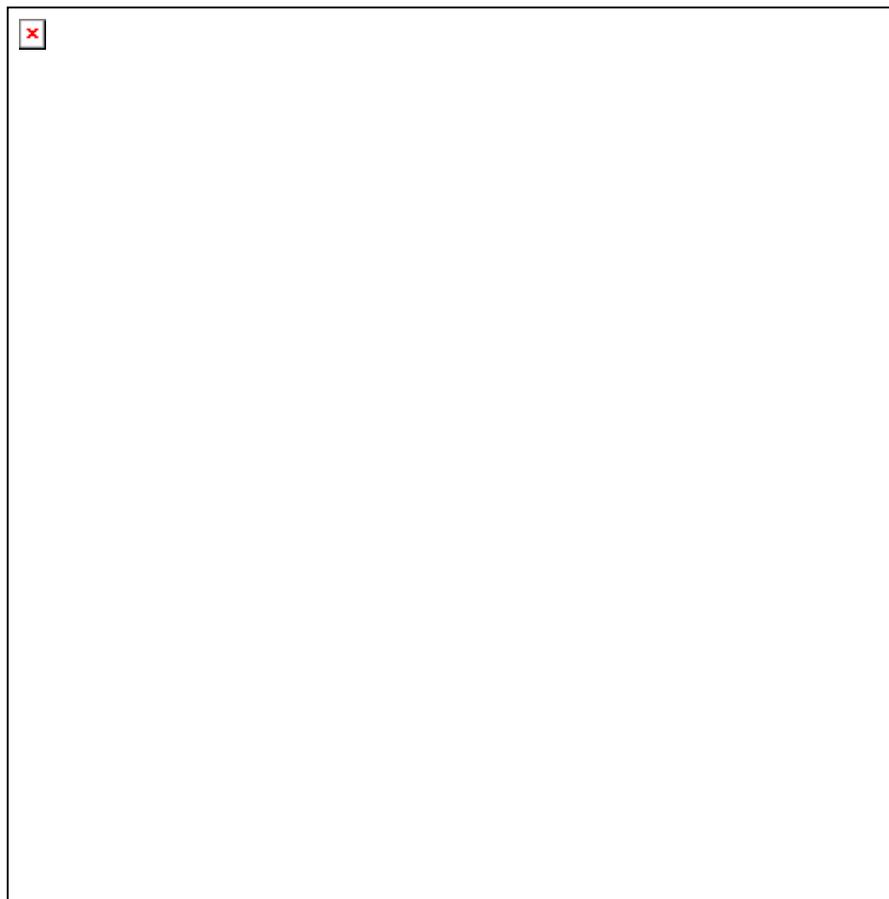


Figure 18. Setting up export to files in the Export Setting dialog box

4.3.7.2. Export to a database

To export to a database, select **Export to ODBC-compatible database** in the **Export type** field (Figure 19).

On the **Connection** tab, enter connection parameters in the **Connection string** window or click **Setup Connection...** and specify connection parameters in the **Data Link Properties** dialog box. Select schema from the drop-down list.

To test the connection with the database, click **Test Connection**.

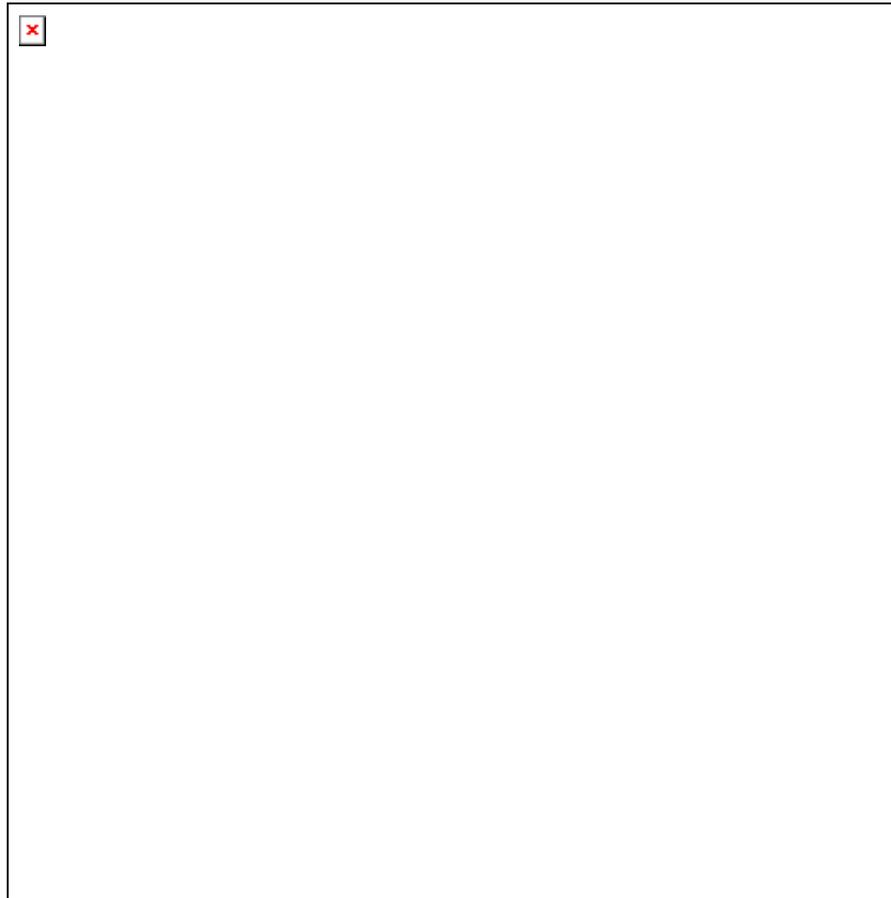


Figure 19. Setting up export to a database in the Export Settings dialog box

Now specify to which tables and table columns of the database the field values of the document are to be exported. To do so, click **Setup Fields Mapping...**

The left-hand part of the **Field Mapping** dialog box (Figure 20) contains a list of document sections and fields. In the right-hand section, specify the corresponding tables and fields of the database. If the database already contains tables for data export, in the **Field Mapping** dialog box select a database table for each section and a database table column for each document field.

If the database does not contain tables for document export, you can create such tables automatically by clicking **Create Tables Automatically**. When you click this button, the program will create tables whose structure is ideal for export. Table rows will be assigned to the corresponding document fields.

Fields groups, fields with multiple instances, document tables, and sections are exported to separate database tables. Two keys are used to link the parent table with child tables: the Primary key and the Foreign key. In the parent table, a Primary key is assigned to each entry, while in a child table a Foreign key containing the value of the corresponding Primary key is assigned to each entry. The keys are automatically added where necessary. You need only specify to which field the key must be exported.

The **Show linked columns** option is selected by default. If this option is not selected, the lists of available table columns in the right-hand side of the tree will not contain columns whose export has already been set up.



Figure 20. Setting up links between document fields and database fields during export

Parameters concerning the saving of images can be specified on the **Images** tab of the **Export Settings** dialog box.

You can save images to a database or to a file (in which case you will need to specify the relevant folder).

Select the format in which images are to be stored. If you select PDF searchable, the program will not only save the image of the document, it will also perform full-text recognition of the document and save the recognized text to the selected format.

If you wish to change the resolution of initial images, for example to reduce the size of stored data, select the **Change resolution to** option and enter a new resolution.

4.3.7.3. Export to a SharePoint library

ABBYY FlexiCapture 8.0 allows you to export document to a Microsoft™ SharePoint library. As result, the documents will be saved in the SharePoint library, and columns with the values from certain fields will correspond to each document. These values can be used for search and indexing purposes.

Notes:

1. You must have administrator rights to be able to set up export to SharePoint. Contributor rights are sufficient for exporting documents.

2. The SharePoint columns into which data are to be written must be of type *Single line of text* or *Multiple lines of text*.

To set up export to SharePoint, select **Export to SharePoint** in the **Export type** field (Figure 21).

On the **SharePoint Connection** tab, type the address of the server (**server URL**) where your SharePoint libraries are located. Use the **Connection settings...** button to set up authentication parameters (Windows logon parameters are used by default) and select Proxy settings if required.

Click the **Connect** button to re-establish a broken connection with the server.

Select **SharePoint library** from the list.

Select the required type of document content from the **Content type** list (starting from SharePoint 2007). In this case, you will be able to export the values into the fields related to the selected content type. Click the **Setup fields mapping...** button and select the required options in the **Field Mapping** dialog box.

Specify the file naming rules: click the **File Naming Options...** button and select the desired options in the dialog box that appears.

On the **Images** tab, specify image storage parameters (file format, image quality, color type, resolution). If you select PDF or PDF/A as an image storage format, you can create searchable PDF files by selecting the **Create searchable PDF** option. You can select a language (or several languages for multilingual documents) different from the language specified in the template: click the **Select...** button to the right of the **Language** field and select the required languages. If you wish to use the language specified in the template, select **As in template** in the **Language** field.

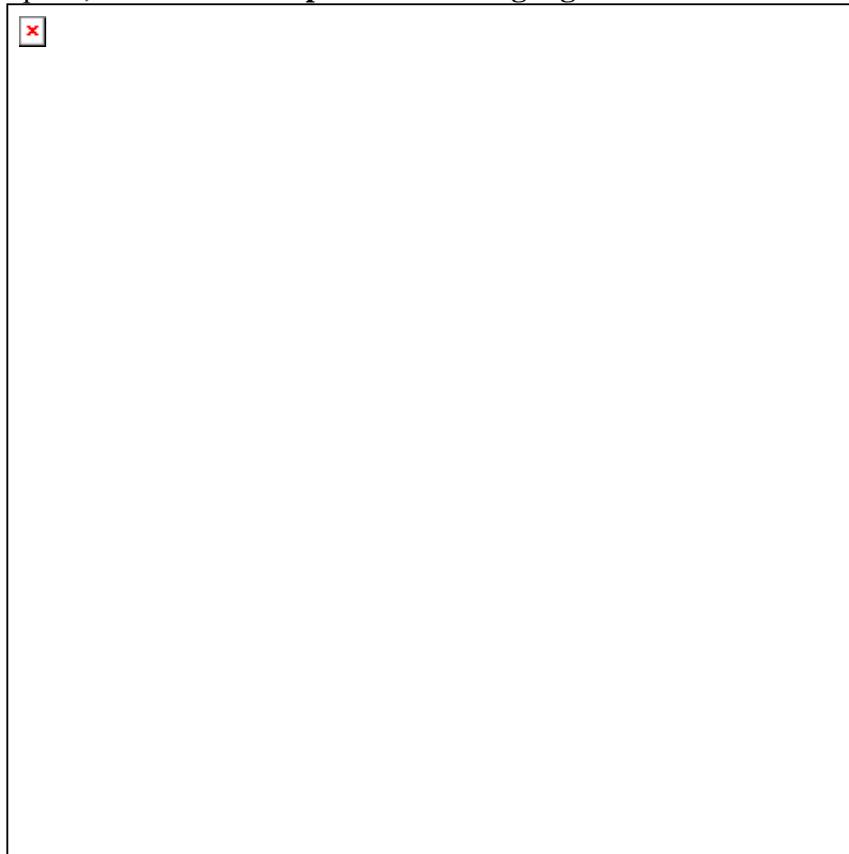


Figure 21. Setting up export to a Microsoft SharePoint library

4.3.7.4. Custom export

This export type allows you to set up advanced export procedures using tools that are not available in the program interface.

If you wish to set up script-based export, select **Custom export (script)** in the **Export type** field (Figure 22).

Next, select the scripting language (**JScript®** or **VBScript**) and enter the script text in the editor window that opens when you click **Edit Script...** (For a detailed description and samples of using scripts, please see the program Help file.)

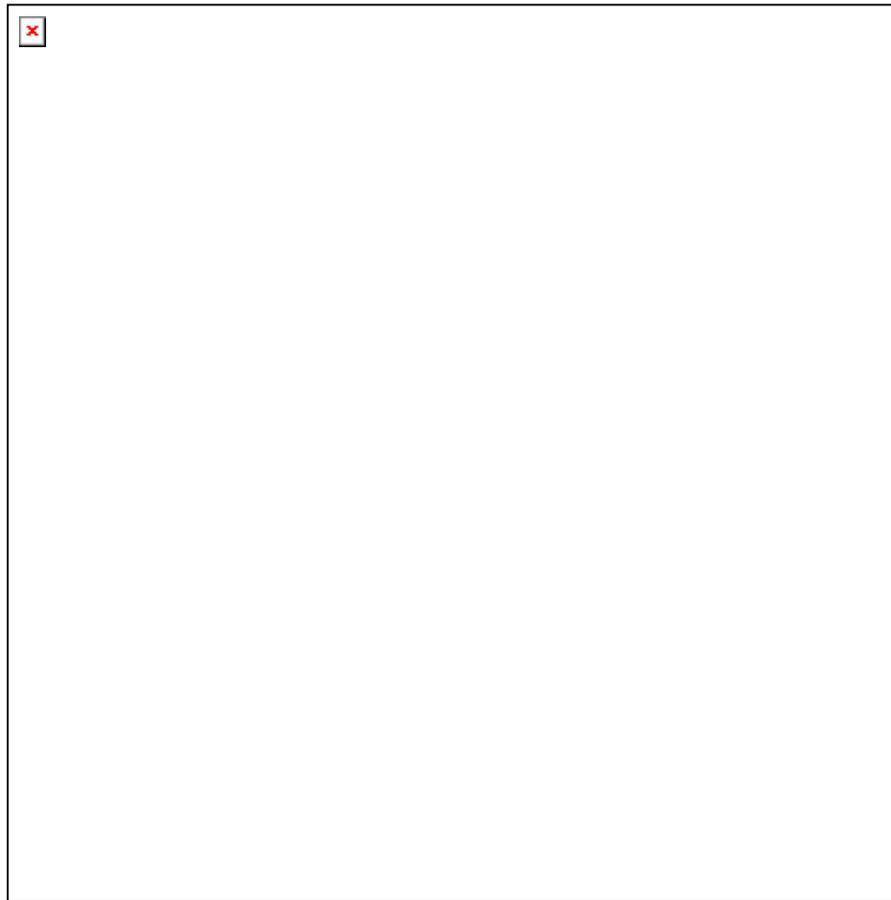


Figure 22. Setting up script-based export in the Export Setting dialog box

Setting up data export options:

1. In the *Document Template Editor* window, select *Template > Export Settings...* In the *Export type* field of the dialog box that opens, select *Export to files*.
2. On the *Destination* tab, enter the path of the folder where the export file is to be saved and specify the name of the export file.
3. On the *Format* tab, select the *XML Document (*.xml)* file type and select the *Errors* option in the *Export additional data* section.
4. If you wish to save page images, go to the *Images* tab, select the *Save document images* option and specify the necessary settings.
5. Once you have specified export settings, the main properties will be displayed on the *Export* tab of the *Document Template Properties* dialog box (*Template>Document Template Properties...*). You can change export settings in this dialog box by clicking *Change Export Settings...*

4.3.8. *Setting up the recognized data view*

Once the data are recognized, the user will see them in the document window. By default, the data are sorted by order and the labels correspond to the template field names. You can however change the view of the recognized data and make it more convenient, for example by changing the order of data presentation or by adding an explanatory text. The recognized data view can be changed in the bottom right-hand side of the Document Template Editor window.

You can move fields, change their names or other properties. To add a label, from the context menu select **Insert Label Box**.

You can specify the font and text size of field names and recognized data values on the **Data Form** tab of the **Document Template Properties** dialog box (**Template > Document Template Properties...**).

See how the recognized data will be displayed in the document window. You can retain the default data presentation or change it.

This step concludes template creation.

4.3.9. *Editing and publishing a template*

Once you have created a template, save it and test it on several images. If you are happy with the result, publish the template so that it is available for document recognition.

To publish a template, click **Publish** from the **Document Templates** dialog box (**Project > Document Templates...** in the main program window).

To return to template editing, in the **Document Templates** dialog box select the template and click **Edit**. The template will be unavailable for editing by other users, and the recognition function will use the latest published version of the template. Template editing will therefore not hinder the work of operators. Once you have edited the template, publish it. The new version will become immediately available for all users. If you do not wish to publish the new version, you can discard the changes and return to the latest published version of the template. To do so, select the template in the **Document Templates** dialog box and click **Discard Changes**.

4.1.10. *Testing the template*

Before you start processing your documents, be sure to test the template.

You can test the template directly in the template editor, on the image that was used to create the template. To start testing the template, select **Document>Run Test**.

Further testing is performed on documents that are added to the test batches, as these batches use the local version of the template. The test batches can be accessed from **Project>Test Batches List** in the main window. You can also access these batches from the template editor by selecting **Tools>Switch to Test Batch**.

If the program detects any rule errors during testing, invalid field properties, etc., you will need to edit the template to correct them. Once all the errors are corrected, publish the template. Now you can start processing your documents.

First, test your template on the source image (**Document>Run Test**), then on the filled-out image of your questionnaire (see [Working with a Set-Up Project](#)). Make any changes that may be necessary and publish the template.

4.4. *Setting up image import*

The operator's first job is to add new images to the project. These images can be paper documents (these must be scanned) or electronic images. If images are regularly received from one and the same source, you can automate the image-adding procedure so that all necessary actions are performed automatically with just a single click. If the necessary settings are specified, the program will scan or import from the specified source and the received images will be processed (for example, despeckled). If necessary, the operator can change the source of image import or change its settings.

If import sources change often and you do not wish to change source settings every time, you can create several import profiles and switch between these profiles.

You can also set up regular image adding in background mode.

To add an import profile, select **Project > Image Import Profiles...** in the main program window. In the dialog box that opens, you can create new import profiles, and edit, delete, or copy existing profiles (Figure 23).

To create a new import profile, click **New**. The Profile Creation Wizard starts.

In the first step of the Wizard, select the import source. A source may be a scanner or a Hot Folder (i.e. a folder where the program will look for new images). If you wish to add images from a Hot Folder, please ensure you have read and write permission for this folder. If the import source is a scanner, make sure that the scanner is connected to your PC.

Creating an import profile using a scanner

1. In the first step of the Wizard, select **Scanner** as the import source.
2. Next, specify the options for processing received images. For example, you can specify that images must be despeckled (necessary if the image quality is low and there is a lot of noise on the image). You can also change the scanning options. In the **Style of the settings dialog** drop-down list, you can select the method of interaction between the program and the scanner: the **FlexiCapture Scanner Settings** method which uses ABBYY FlexiCapture dialog box for setting up scanning options, or the **Scanning Options** method which uses the dialog box of the corresponding scanner driver to set up scanning options.
3. Next, specify document assembly options and image pre-processing options. For example, you may set up the program to despeckle images (this may be necessary for noisy images), to convert images to black and white, to rotate images in a certain direction, or to disable skew correction.
4. Finally, change the name assigned by default to the import profile and enter a description.

Creating an import profile using a Hot Folder

1. In the Wizard, select **Hot Folder** as the import source and specify the path to the Hot Folder.
2. Next, specify the image loading options. You can select the **Check hot folder every** option and specify the interval following which the program will check the Hot Folder for new images. In the **Batch Settings** section, select one of the options to specify the batch to which new images will be added. In the **Number of files to add** drop-down list, select one of the values to specify the number of files to be added to the selected or newly created batch.
3. Next, similarly to creating an import profile for the scanner, you can specify document assembly options and image pre-processing options.

4. The Import Profile Creation Wizard now prompts you to set up options for purging the Hot Folder following import. Images that were successfully imported and images whose processing produced an error can be deleted or moved to another folder.
5. Finally, change the name assigned by default to the import profile and enter a description.

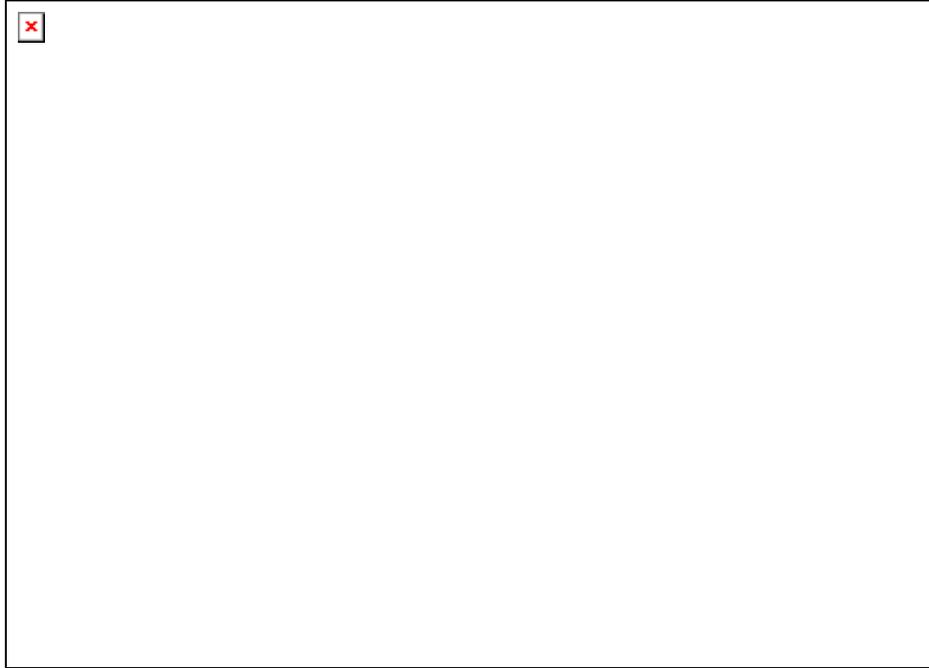


Figure 23. The Image Import Profiles dialog box

1. Set up an import profile for your images. To do so, select *Project > Image Import Profiles* in the ABBYY FlexiCapture 8.0 Professional main window.
2. In the Image Import Profiles dialog box, click *New...* to create a new profile. Enter the profile name and select the *Scanner* option (the questionnaire images will be received from the scanner).
3. Set up scanning and image processing options.
4. On the last step of the wizard work you can change the default name assigned to the import profile and provide a description.

5. Setting Up the System for Processing Flexible Documents

Setting up the processing of flexible documents is a more difficult task. Since the location, size, and number of data fields vary between different documents, you cannot create a template with fixed arrangement of fields for invoices, payment orders, and other similar documents. Therefore a fixed template cannot be successfully matched to all copies of such documents. On such documents, data fields are searched for after the page is recognized as a whole, and the template is created on the basis of keywords and mutual arrangement of data fields. A special component of the system, ABBYY FlexiLayout Studio 8.0, is specifically designed for creating such templates. For a detailed guide on creating a flexible template, please see the ABBYY FlexiLayout Studio 8.0 Help file.

A created flexible template is exported to an AFL file and then attached to ABBYY FlexiCapture 8.0 Professional. We can use a flexible description for the entire template or as one of the sections. A flexible description can be multi-page.

You can attach a flexible description at the stage of document template creation. To do this, add the document image during the second stage of template creation. Next, select the **Load FlexiLayout** option and enter the path to the AFL file containing the flexible description.

You can also attach a flexible description in the Document Template Editor window. To do so, use the Properties dialog box of a section. To open the **Properties** dialog box of a section, from the context menu of the section select **Properties...** Open the **FlexiLayout** tab and click **Load...** Select the AFL file containing the necessary flexible description.

Once the flexible description is attached, all fields and their marking will already be present on the image.

The set of fields in the template must be identical to the set of fields in the flexible description. You must not add or remove fields (except fields with no regions on the image). You can change the set of fields to be recognized only if you change the flexible description.

Next, specify the properties of data fields. This process is similar to setting up field properties for a fixed template. Set up recognition, verification, and export properties and specify the necessary rules.

For more information, please see [Creating a document template](#).

6. Peculiarities of Capturing Non-Structured Documents

ABBYY FlexiCapture 8.0 Professional can help you process non-structured documents containing information written in the free-form style, for example contracts, letters, order, appendices, etc. Non-structured documents with text or images separated by blank pages or pages with barcodes can be processed using ABBYY FlexiCapture 8.0 Professional and exported to searchable PDF files or files in graphical format.

Processing such documents usually means converting them to electronic form and organizing the search based on key fields.

The key fields on such documents are usually searched for with the help of a flexible description created in ABBYY FlexiLayout Studio.

If automated search for key fields is unavailable, the operator can enter the values for the key fields manually. To do so, create a document template with a single field (or multiple fields, if necessary) and select the **Don't recognize (Key From Image field - will be entered manually)** option on the recognition properties for this field or fields. When you start the verification process, the operator will now be prompted to enter the value of the key field (or fields) manually.

To store the document, you must set up the export process. The key field values can be exported to a file or to a database and the document image can be saved in your preferred graphical format. To do so, select the **Save document images** option and specify the image saving parameters on the **Images** tab of the **Export Settings** dialog box. You can save document images as searchable PDF files or files of a selected graphical format.

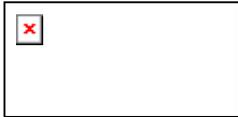
Pay close attention when assembling pages into documents: with unstructured documents it can be fixed to determine to which document a particular page belongs. To automate the assembly of unstructured documents we recommend separating documents with blank sheets or pages with barcodes. When adding images to a batch (by scanning, adding from file, or creating an import profile) you then need to

enable the option **Images separated by** and select the value **blank pages** or **pages with barcode** from the dropdown list, depending on which pages are to be used as separators. Pages are assembled into documents automatically: pages will be added to the current document until the next separator page.

7. Working with a Set-Up Project

Once the administrator has set up a template and specified all the necessary settings, you can start processing documents. The entire process can be set up so that the operations of adding images, recognition and data export are performed with a minimum of human intervention. The work of the operator does not depend on the document type.

Document processing consists of four stages. Each stage has a corresponding button on the toolbar.

<p>1. Import</p> <p>Click on the arrow to select one of the image import options: Load Images..., Scan Images, or import based on one of the specified profiles. The button text corresponds to the performed action.</p>	
<p>2. Recognition</p> <p>Starts the process of document recognition. By clicking on the arrow you can also run the document analysis or match one of the templates.</p>	
<p>3. Verification</p> <p>Starts the verification process. By clicking on the arrow you can also start the process of rules check.</p>	
<p>4. Export</p> <p>Starts the data export according to the properties specified in the template. By clicking on the arrow you can export data to a file or database.</p>	

7.1. Adding images

The first stage of processing documents is to add page images to the project. To begin working, select or create a batch to which images will be added (to create a batch, right-click in the program main window and select **New Batch**). If you try to add images to a project which has no batches, a batch will be created automatically.

There are several methods by which images can be added to a batch:

1. Load previously saved images from files. To do so, select **Load Images...** from the drop-down menu of the import button or press Ctrl+O. In the dialog box that opens, select the necessary image files and set up the image import options.

When importing images from multi-page files, multiple pages will be added to the batch.

2. Scan images. To add images from a scanner, from the menu select **Scan Images...** You will be prompted to select a scanner and scan the images.
3. Import images with the help of an image import profile already created by the administrator (see. [Setting up image import](#))

If import profiles have been set up, their names will be displayed in the menu of the import button. To start importing images, select the name of one of the import profiles.

When you select **Import Images...**, the **Select Import Profile** dialog box opens. In this dialog box, select an import profile from the drop-down list and click **Import** to start importing images. In this dialog box you can also create a new image profile or edit existing profiles..)

If import profiles have been set up, their names will be displayed in the menu of the import button. To start importing images, select the name of one of the import profiles.

When you select **Import Images...**, the **Select Import Profile** dialog box will open. In this dialog box, select an import profile from the drop-down list and click **Import** to start importing images. In this dialog box you can also create a new image profile or edit existing profiles.

If you have selected an import profile once, you no longer have to search for it in the list when you need to use this profile again. Instead, just click on the import button where the name of the recently used profile is displayed.

Images can be imported in background mode if a corresponding import profile is set up. Images from the Hot Folder are imported into the system automatically. The program checks the Hot Folder for new images after the time interval specified in the settings.

Once new images are added to a batch, unprocessed pages will appear in the list.

1. Create a batch to which the image will be added. To do so, right-click in the program main window and select *New Batch*. Enter the name for the new batch and its description. Double-click on the batch to open it.
2. Fill out the questionnaire.
3. Scan the image.
4. Import the image by selecting *Import Images...* or *Scan Images...* A page with the <Unknown> name will appear in the batch.

7.2. Recognition

To recognize data, click **Recognize**. The program matches the template, selecting it on the basis of anchors and identifiers, and recognizes data in the document regions specified by the fields of the selected template.

Images can be recognized automatically and immediately after the images have been added to the batch. To do so, select the **Recognize added images automatically** option (**Tools > Options... > Document Processing** tab).

The **Confidence Level** displays the percentage of reliably recognized characters relative to the total number of characters. Once recognition is complete, the operator can move to the document verification.

Prior to recognition, you can analyze a page, i.e. match the template with this page. To do so, select **Analyze** on the drop-down menu of the recognition button. If the template can be matched, field names

appear for the page and for the document to which the page belongs, and field recognition is performed. If none of the templates of the project can be matched with the page, the page remains unprocessed.

In most cases a correctly created template matches with pages automatically. However, sometimes you may need to select a template manually. To match a template, select the necessary page or document and select **Match Template...** from the drop-down menu of the recognition button.

Select *Analyze* on the drop-down menu of the recognition button. If the template is matched successfully, its name will appear instead of <Unknown>.

Click *Recognize*. The page will be recognized.

7.3. Verification

Verification, i.e. checking recognized data, is the most demanding part of the operator's job. The verification process in ABBYY FlexiCapture 8.0 Professional is therefore organized so as to maximize convenience for the operator and minimize the number of errors. For multi-page documents, the program first checks that pages have been correctly assembled into documents. The verification process then starts. This consists of group and context verification. You can also run a check in the document window. Rules are checked during the verification process.

Document assembly check. For multi-page documents, the program checks that pages have been correctly assembled into documents. If the order of pages is not identical to the specified order or if the values of the key field are not identical on all pages, the document is marked with a red flag and the error message is displayed in the document window. The operator must make sure that the pages were not mixed up during scanning. Sometimes assembly errors can be corrected simply by changing the order of pages.

A convenient way of checking the assembly of pages into documents is by using page thumbnails mode (Figure 24). Here you can change the order of pages or even move pages from one document to another using drag-and-drop.

If the template specifies that a key field check must be performed for correct document assembly, the values of key fields will be displayed below the image of each page. If the key fields on the pages of a document are not identical, they will be displayed in red. Key fields may differ because they were recognized or filled out incorrectly. Please check the values of key fields. If these values are still not identical, the pages probably belong to different documents. If the order of pages has been mixed up, locate pages with identical key fields and assemble them into documents.

Note. To zoom in on the details of pages thumbnails, hold down the Ctrl key and use the mouse wheel.



Figure 24. The program main window, page thumbnails mode

To begin verifying the recognized data, click **Run Verification...**

Group verification means grouping character images which have been recognized as having an identical value and displaying them on the verification screen in order to confirm correctly recognized characters and leave for the next stage only those characters which are either incorrect or uncertain (Figure 25).

During group verification, the operator can view the image of the field where the checked character is found. To do so, select **Show Character Image** from the shortcut menu of the selected character or press F2. By positioning your cursor over the checked character, you can also activate the mode which will display the field where the character is found. To do so, in the verification window, select **View > Field Image > Show Field Image** or press Ctrl+I.

To correct incorrectly recognized characters, proceed as follows: select a character that does not correspond to the group character and enter its correct value. The entered value will be displayed in green in the upper left-hand corner of the character image. If you are not sure of the value of the character even after you have viewed its context, left-click the character to mark it with a red interrogatory mark. You can also change the character status using the button on the **Toggle** toolbar.

To confirm a correctly recognized character, select **Confirm** from the context menu. Alternatively, on the toolbar click **Confirm All** to confirm all displayed characters at once.

During template creation, you set up verification options when specifying the field properties. Group verification is performed for characters from fields for which you have selected the **Include in group verification** option on the **Verification** tab of the **Properties** dialog box.

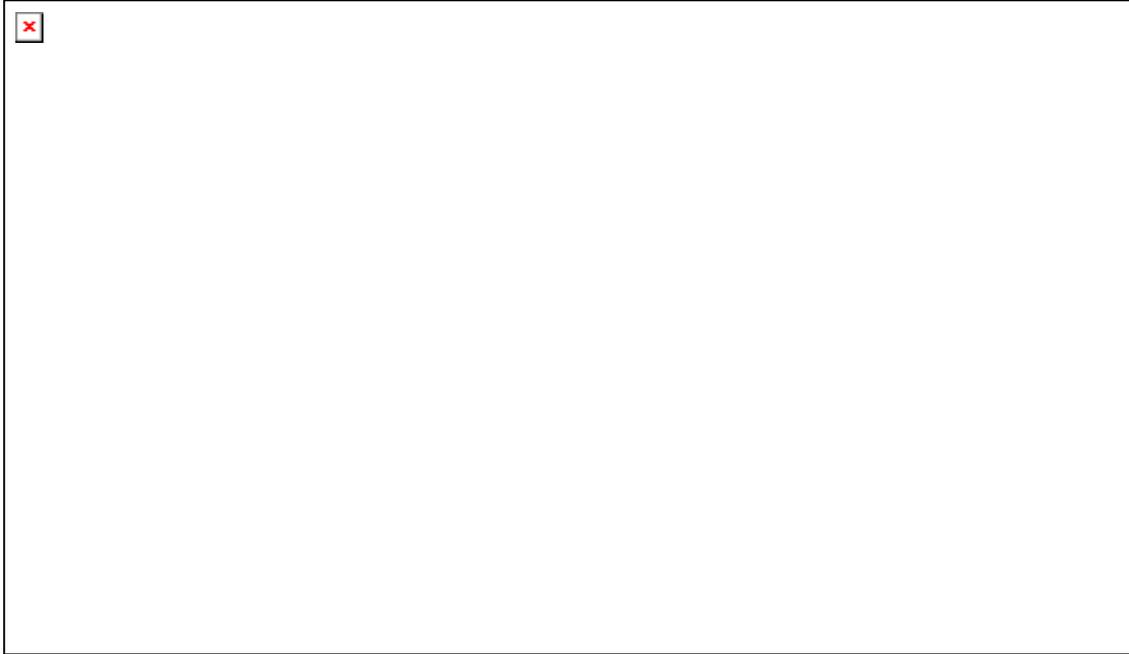


Figure 25. Group verification of digits

Context verification is a verification mode used to correct the format of fields whose value range is known or easily identified. Country name is one such field because we know what values this field can have.

To make a decision about the necessary changes in the field, the user must only see the field itself and know the set of its possible values. Both fill-out errors and field format errors can be corrected.

Context verification is performed for fields for which you have selected the **Include in field verification** option on the **Verification** tab of the **Properties** dialog box.

To correct incorrectly recognized characters, use standard text editor modes, for example insert mode and overwrite mode. To switch between modes, press **Insert**.

Users view the recognition result of each field in turn, correct it, and confirm it by pressing **Enter** or clicking **Confirm Field**.

Fields that do not correspond to the specified data type are marked with a red flag and an error message is displayed for such fields. Fields for which you have specified rules but whose values do not meet the requirements of these rules are similarly marked. You must correct the values of such fields. If you cannot do so, you must postpone recognition of the field value by clicking **Postpone**.



Figure 26. The field verification window

Document window also allows you to check that recognition was correct and to correct erroneous characters (Figure 27). The document window opens when you double-click on the name of the page. This window consists of the data area, page image, and rule errors area (if there are such errors). You can set up the arrangement of windows with the help of the **Layout** button. In the document window the operator can see the entire document, not just groups of characters or separate fields.

In the data area, uncertain characters are marked in red, and fields with wrong data types or rule errors are highlighted. To switch to the previous or next error, use the  and  buttons. These buttons allow you to scroll through assembly errors, uncertain characters, rule errors, etc.

Fields for viewing and editing can be arranged in sequential order or in any order convenient for the user. Their arrangement can be changed in the Template Editor, see: [Setting up the recognized data view](#).



Figure 27. Document window

Rules check. Rules whose requirements are not met are marked with either a yellow flag (warning) or a red flag (error). If a rule relates to one of the fields, such a field must be sent to the verification operator during context verification. Rule errors are displayed in a separate window of the document editor, and documents which do not meet rule requirements are indicated by red flags.

If the requirements of a given rule are not met, the operator must check that data were correctly recognized and correct any recognition errors. If the error is a fill-out error that cannot be corrected, the operator must not export the document.

To start re-checking rules errors, click on the arrow to the right of the button and select **Re-check Rules**. If the requirement of the rules are met after you have corrected the field values, the flags will be removed.

1. Click *Run Verification...* and verify the recognized data.
2. See which characters and fields are sent to group and context verification. Verify them.
3. If any fields are marked with a red flag and an error message, ensure that the values of such fields correspond to the specified data type.
4. If, for example, the fill-in date does not correspond to the specified format or does not belong to the specified time interval, please ensure that the data in the Fill in date field have been recognized correctly. Once you have made changes to the data, the error message may disappear from the corresponding area.

7.4. Export

Once the recognized data have been verified, the operator exports the batch by clicking **Export**. The export is performed according to the document template settings.

If you do not want to use the document template settings, you can export the data to a file or database by selecting corresponding items in the drop-down menu of the **Export** button. When doing this, you can specify any export settings.

Export data to a file according to the previously specified settings. To do so, click **Export**. Open the resulting file and analyze the export results. Send this file to ABBYY by e-mail. The address is FlexiCapture_Feedback@abbyy.com.

Thank you for your help and may we wish you every success with our software.

8. Conclusion

This simple example covered all the stages of program set-up and processing structured documents. The capabilities of the program, however, are much greater. It can help you process simple and complex multi-page documents of various types: semi-structured, non-structured, and mixed-type documents. If you have any questions, please refer to the program Help files and to the Installation Guide.