

**System of automatic input of forms
and questionnaires**

FineReader 4.0 Forms

User's Guide

**ABBYY Software House (BIT Software)
Moscow 1998**

ABBYY Software House (BIT Software)

FineReader 4.0 Forms

User's Guide

Information in this document is subject to change without notice and does not represent any commitment on the part of ABBYY (BIT Software).

The software described in this document is supplied under a license agreement. The software may be used or copied only in accordance with the terms of the agreement. It is against the law of the Russian Federation "On legal protection of software and databases" and the standards of international law to copy the software onto any medium except as specifically allowed in the license or nondisclosure agreements.

No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without the express written permission of ABBYY Software House.

© ABBYY Software House (BIT Software), 1993-1998. All rights reserved.

ABBYY, BIT Software, FineReader, "fontain image transformation", Lingvo, Scan&Read, Scan&Translate, "one button principle", "Your computer reads by itself", "Your computer reads and translates by itself" are registered trademarks of ABBYY. Try&Buy, DOCFLOW are trademarks of ABBYY. All other marks are trademarks or registered trademarks or their legal owners.

105568 Moscow, box 19. ABBYY

Contents

License Agreement	7
Part 1. Operator's guide: work with already tuned program.....	10
Program start	10
Main window of the program	10
Batch window	10
Image window and Form window	11
Scan&Read toolbar	11
Standard toolbar.....	13
Adding pages to the batch	13
Scanning	13
Adding pages from image files	13
Selecting pages	14
Deleting whole pages or their components	14
Template matching and recognition	15
"Read and wait mode"	15
Recognition in the background mode.....	16
Scanning and recognition with one button	16
Editing recognition results	17
How the recognized form looks like.....	17
Checking recognition results	17
Working with the dialog of results checking	18
Checking with validation rules	20
Navigating in the recognition results	20
Typical problems when editing	21
Export.....	22
Part 2. Administrator's guide: Tuning the program	25
Software and hardware requirements	25
Installing FineReader	25
Installing and tuning the scanner	26

Installing the scanner	26
How the scanner works with FineReader	26
Assembling a batch from image files	26
If you scanner doesn't work with FineReader	27
Setting scanning parameters	27
Main scanning parameters	28
Scanning multiple pages	28
Scanning different types of forms	29
Numbering pages in the batch	29
Machine-readable forms	30
What does machine-readable form (MF) mean	30
General requirements for MF	31
Requirements for reference blocks	31
Requirements for descriptive information	32
How the customer fills the form in	33
Auxiliary notes	33
Certification	34
General scheme of tuning the system	34
Creating a batch	34
The algorithm of batch creation	34
Setting batch options	35
Creating template	36
Algorithm of template creation	37
Preliminary editing of the form template image	37
Marking out fields on the template image	37
Field types on the template image	37
Selecting fields on the form template	38
Deleting fields	39
Text field parameters	39
Default parameters	40
Individual parameters for text field	40
General, Advanced and Checking options	41
Control fields	42
Additional fields	42
Manual input fields	43
Setting export options	43
Export to database	43
Export to file	44
Export via Automation	45

What you get after export	45
Multiple templates	46
For better distinguishing of templates	46
User's languages of recognition.....	47
When to create a new language.....	47
Creating user's languages and language groups.....	47
Characters for the user's language or user's group of languages	48
Attaching a dictionary to the user's language	48
Checking recognition results	50
Validation rules to check recognition results.....	51
Validation rules list	51
How to create a rule	51
Automation-check.....	52
Replacing values from the list	52
Replacing characters from the list.....	54
Normalization and check of date	55
Normalization of price.....	56
Merging the fields	56
Checking against database	57
Check with expression	58
Sum checking	58
Sum in digits – sum in writing.....	59
Regular expression	59
Training a pattern	61
What does to teach a program mean.....	61
How the system is trained	61
Batch processing in the network	62
How to work with a batch in the network	62
What is FineReader Automation API	62
Part 3. Appendices	64
Supported languages of recognition	64
Handprint recognition	64
Printed recognition	64
Input image formats	66
Hot keys	67
ODBC Drivers	70

License Agreement

ATTENTION! This SOFTWARE (including the media and printed materials) is transferred to you on the terms of End-User License Agreement (EULA), NOT SOLD.

Before opening the packet, study the End-User License Agreement (EULA). The opening of the packet means that you agree with the terms of the EULA.

If you do not agree with any terms of this EULA, then within seven days of buying the SOFTWARE return the unopened SOFTWARE (with media and documentation) to the company from which you obtained it.

END-USER LICENSE AGREEMENT FOR THE SOFTWARE

The End-User License Agreement (EULA) for SOFTWARE is the legal agreement between you (either an individual or a single entity) and ABBYY Software House.

The enclosed document "Proof of Purchase" is part and parcel of this EULA.

The software ("SOFTWARE") distributed by ABBYY is a set of programs, databases and documentation, protected by copyright laws. For the purposes of this text, "documentation" means printed material and text files, carrying the description of the SOFTWARE components, which are the integral part of the SOFTWARE.

1. Subject of EULA

1.1 The subject of this EULA for SOFTWARE is the non-exclusive license on this SOFTWARE use, transferred by ABBYY to the End-User.

1.2 All items listed below refer both to the SOFTWARE in the whole and to its components in particular.

2. Copyright

2.1 Exclusive copyright to this SOFTWARE is owned by ABBYY Software House.

2.2 As the End-User you get a license for the use of the SOFTWARE if you accept the following terms and conditions.

2.3 The license is given exclusively to you (unless there is additional special written permission on the part of ABBYY Software House).

3. Terms of the SOFTWARE use

3.1 You may use the SOFTWARE at one and the same time only on the number of processors specified in the "Number of licenses" column in the "Proof of Purchase". You may not use the SOFTWARE on a larger number of processors at one and the same time. You may not store (whether installed or not) more copies of the SOFTWARE on than you hold licenses.

3.2 You may make one backup copy to retain as your SOFTWARE Master Backup Copy for use only should the original be lost or damaged. The backup copy may not be used for any other purpose.

3.3 You may create your own databases for programs which are part of the SOFTWARE if such possibility is mentioned in the documentation.

3.4 If you receive the SOFTWARE on more than one medium, regardless of the type or size of medium you receive, you may use only one medium that is appropriate to your single computer. Both media are considered to contain the same copy of the SOFTWARE.

3.5 You may not redistribute the SOFTWARE. Under redistribution of the SOFTWARE we mean giving access to the third parties to the SOFTWARE components copied by any company, by network and other ways, or sold, rent or leased.

3.6 You have limitations on the following activities:

3.6.1 You may install and use the Software on, at most, one processor at any given time. If the software is to be installed on a network, it may be accessed by only one workstation at a time;

3.6.2 You may not reverse engineer, decompile or disassemble programs, databases and other SOFTWARE components;

3.6.3 You may not make any changes in the object code of the program or of databases except those specially provided by the SOFTWARE and described in the documentation;

3.6.4 You may not give copyright on the use of the SOFTWARE or any other rights on the SOFTWARE to third parties;

3.6.5 You may not do any other actions that violate Russian and international laws on copyright and software use.

4. Termination

4.1 This EULA for the SOFTWARE comes into effect after the SOFTWARE is installed on the computer and remains in force during the time you use the SOFTWARE.

4.2 If you fail to comply with the terms and conditions of this EULA, you must destroy all copies of the SOFTWARE (including printed materials, media (disks), information files, archive copies of the SOFTWARE), and the agreement is terminated.

5. Responsibility

5.1 Illegal use, redistribution and copying of the SOFTWARE is the breach of the law of Russian Federation "On the legal software protection" and is sued according to the Law.

5.2 If you fail to comply with the terms and conditions of this EULA, ABBYY immediately withdraws your right to use the SOFTWARE and cancels its warranties, technical support and free delivery of the SOFTWARE.

6. Warranty

6.1 ABBYY guarantees the quality of data on media, working capacity of programs enclosed in the SOFTWARE packet under conditions described in the documentation, conformation of the SOFTWARE components to the specifications, and the typographical quality of documentation.

6.2 As for the other components, the SOFTWARE is delivered "as is". ABBYY does not warrant that the software carries no errors, nor does it take on any liability for consequential damages, either direct or indirect, including damages caused by possible errors or misprints in the SOFTWARE package.

6.3 ABBYY makes no guarantees as to the operation of this software in combination of any other installed software or hardware, especially any models of software or hardware of earlier release.

6.4 The restricted warranty remains in force during 60 days from the dated of the SOFTWARE purchase. During this time ABBYY takes all claims to the SOFTWARE quality.

ABBYY (BIT Software, Inc.).

105568, Russia, Moscow, Chechulina 20-1.

For letters: 105568, Russia, Moscow, PO box 19

Phone no.: +7 095 263-6658, 263-6659.

Fax: +7 095 263 6278.

E-mail: sales@abbyy.ru

<http://www.abbyy.ru>

How to use this guide

This is the guide's guide for the is used for **FineReader Forms** program (versions FineReader Professional and FineReader Handprint). The guide describes the entire set of features supported in FineReader Handprint. As for FineReader Professional, it lacks the following features:

- Recognition of handprinted characters
- Check of recognition results
- Check of recognition results with validation rules
- Multiple templates
- Export to databases via ODBC and Automation

The user's guide consists of two parts. Part 1, "Operator's guide: work with already tuned program ", is the guide for an operator, who inputs forms (scans, recognizes, edits and exports them). The operator works with already tuned program (with ready batches and ready templates).

Part 2, "Administrator's guide: Tuning the program" describes batch creation and tuning.

Appendices, which will be of use to administrator, contain additional information about program features.

P a r t 1

Operator's guide: work with already tuned program

FineReader Forms is a powerful tool for bulk input of machine-readable forms in information systems and databases (for details on machine-readable forms see "Machine-readable forms", p. 30). Its working principle is very simple: you prepare everything for scanning and click the *Scan&Read* button. The program gets the image of the current document from the scanner and marks out fields for recognition, automatically assigns them to the corresponding database fields and recognizes them. The recognition result is sent to a file or database.

Program start

Attention! For the normal work of the program you need a hard-lock, without which the program works only in the demo mode.

To run FineReader Forms:

- Select *FineReader 4.0 Handprint Forms* from the *Start/Programs/ABBYY FineReader* menu.

Main window of the program

To open a ready batch:

1. From the *File* menu, select *Open...*
2. In the displayed dialog select the necessary file with the .frm extension. You will see the Batch window.

Batch window

Batch is a set of documents. It contains images and/or recognized text. In the *Batch* window you can see the list of pages of the current batch. To view a page, double-click its icon or number. The files of the page will open: text and/or image, — in the Text and/or Image window, respectively.

Columns in the Batch window

In the Batch window (view with parameters) you can see the following columns:

Column title:	Meaning:
Number	Page number in the batch.
Opened by	Name of the user (and computer), who opened the batch page.
Image	Whether the page has an image.
Template	Name of template with which the page is recognized.
Text	Whether the page has text (whether it is recognized).
Checked	Whether the page is checked (no unchecked characters on the page).
Export	Whether the page was exported.
Error count	Number of recognition errors (including indefinitely recognized characters) on the page.
Error	Whether there is an error during page processing.
Rules	Whether there was an error in validation rules on the current page.
Edited by	Name of the user (and computer), who edited the page.
Comments	Additional information about the page.

Image window and Form window

In the Image window you can see the image of the current page. The Form displays the recognition results of the current page and the list of rule errors (if any).

Scan&Read toolbar

The Scan&Read toolbar (Figure 2, p. 12) contains buttons that carry out the main operations of the program. The numbers on the buttons show the order of operations carried out to get an electronic version of the paper document: 1 – Scanning, 2 – Matching the template, 3 – Recognition, 4 – Checking results (rules

checking and recognition results checking) and 5 – Export (to file, to database, via Automation).

The leftmost button *Scan/Open&Read* can work in the Scan&Read mode (for paper documents) or in the Open&Read mode (to add electronic images to the batch).

To know the mode of the button on the Scan&Read toolbar, position the mouse pointer over the button. You will see a tooltip under the button (Figure 1).

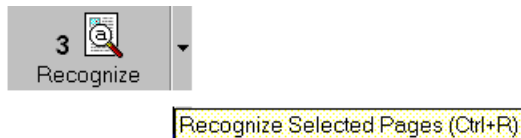


Figure 1. Tooltip for the “Recognize” button (Scan&Read toolbar).

To change the button mode, click the arrow to the right of the button: you will see the menu (Figure 2). To set another mode, select the corresponding menu item.

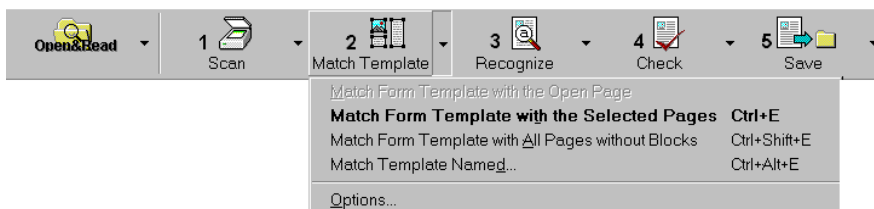


Figure 2. Open menu of the “Match template” button.

The mode of the *Match template* and *Recognize* buttons depends on what window is active and also whether the *Batch* window has any pages selected. If the *Image* or *Form* window is active, the operation is carried out for the open page. If the *Batch* window is open, then these buttons (*Match template* and *Recognize*) work with *selected pages*.

To save screen space, you can make the Scan&Read toolbar buttons small (*Tools menu> Options...> General tab> Show group> Large buttons on the Scan&Read toolbar*).

This time the Scan&Read toolbar will look like the following:

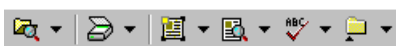


Figure 3. Scan&Read toolbar: small buttons view.

Standard toolbar

The Standard toolbar has the following buttons: seven standard buttons (new, open, cut, copy, paste, undo, redo), buttons to search and move in the text, buttons to move to the next and previous pages, buttons to rotate and clean images. On the Standard toolbar you can also see the list for the image scale, the button to move all blocks in the Image window and the *Help* button.

The buttons for image rotation and cleaning work with an open image or with selected images.

The *Move blocks* button becomes active after the template is matched with the image.

When you click the *Help* button, the mouse pointer takes a special shape (an arrow with a question mark). If you click the mouse on any object (button or menu item), you will get its brief description.



Figure 4. Standard toolbar.

Adding pages to the batch

To add pages to the batch, you can scan paper images or open ready image files.

Scanning

To scan one pages:

- Click *I-Scan*.


To scan multiple pages:

- Click the arrow to the right of the *I-Scan* button and from the open menu select *Scan multiple pages*.

Adding pages from image files

You will find the list of input formats in Appendices, p. 66.

To add an image file to the batch:

1. Click .
2. In the *Open* dialog you can select one or more images to add them to the batch.

Note: In this dialog you can simultaneously select several files. Hold the SHIFT key to select a number of consecutive files, or CTRL – to select non-consecutive files.

You can copy or move the files you add to the batch. To move the files, select the corresponding checkbox in the *Open* dialog.



With one button click you can add a file to the batch and recognize it:

1. Click the arrow to the right of the *Scan/Open&Read* button and from the open menu select *Open and read*.
2. In the *Open* dialog you can select one or more images, which will be added to the batch and immediately recognized.

Selecting pages

The batch supports multiple selection of pages. This means that you can select many consecutive or non-consecutive pages at the same time.

To select a number of consecutive pages, click the mouse on the first page. Holding the SHIFT key, click the last page of the selection.

To select a number of non-consecutive pages, click them with the mouse holding the CTRL key.

To select all pages of the batch, activate the Batch window and from the *Edit* menu choose *Select All...*

Note: 1. What *Select All* command does depends on what window is active – Batch, Image or Form: this command selects pages in the active batch, or all blocks on the image, or all text in the current form field.

2. It is more convenient to select several pages when the Batch window is displayed with parameters (Figure 9, p. 35) (local menu, *Display Page Properties* item).

Deleting whole pages or their components

If you do not need this or that page in the batch, you can delete it. Since the pages in the batch can consist of three files: image – blocks – text, then you can delete either all components of the page or only some of them, e.g., text only, image and blocks only, etc.

To delete the whole pages (image, text and blocks):

- Select the necessary pages and press DEL.

To delete some components of the page:

1. Select the necessary pages and from the *Edit* menu select *Delete Pages...* (hot key is ALT+DEL).
2. In the displayed dialog select the selected pages' components you want to delete and click *OK*.

Template matching and recognition

The form template describes position, sizes and attributes of fields on the page. Before the recognition the form template is matched with the current image. The template is matched with the help of reference blocks (the image shift and the skew up to 10 degrees are compensated for). The template contains information what fields should be recognized.

To match the template with selected pages of the batch:

- Click the *2-Match template* button.

To match template with all batch pages which do not have blocks yet:

- Click the arrow to the right of the *2-Match template* button and from the local menu select *Match form template with all pages without blocks*.

To start recognition of selected pages:

- Click *3-Recognize*.

To start recognition of all unrecognized pages of the batch:

- Click the arrow to the right of the *3-Recognize* button and from the local menu select *Recognize all unrecognized pages*.

If you start recognition of the pages which do not have blocks yet, then the program will firstly match the template with them and only then start the recognition.

The recognition results are displayed in the Form window (see “How the recognized form looks like“, p. 17).

If the template could not be matched or if the recognition results leave much to be desired, see the part “Typical problems when editing“, p. 21.

“Read and wait mode”

The *Read and wait* mode is used to organize an automatic recognition in the network (for details see “Batch processing in the network“, p. 62), since in this mode the recognition is automatically resumed when new unrecognized pages appear in the batch.




To start recognition in the “Read and wait” mode:

- Click the arrow to the right of the *3-Recognize* button and from the displayed menu select *Recognize all unrecognized pages and wait*.

Recognition in the background mode

FineReader allows you to run recognition and edit recognized pages at one and the same time, that considerably speeds up the processing of documents on one and the same computer.*

Start recognition in the background mode (*Scan&Read>Recognize in background mode*). In the Status bar you will see the  indicator. If page parameters are displayed in the batch (local menu, *Display Page Properties* item), then you will see the *Background recognition...* line near the page under recognition. You can also see the bullets that appear opposite the processed pages in the *Blocks* and then – in the *Text* columns.

While the program is recognizing, you can open and edit the recognized pages.

Scanning and recognition with one button

The *Scan&Read* button allows you to scan and immediately read one or multiple pages.

To set the desired working mode, click the arrow to the right of the button and from the local menu select the desired item.

* To recognize pages in the background mode, a second copy of the program is run.

Editing recognition results

How the recognized form looks like

The screenshot shows a form window with the following fields and values:

ABBYY	ABBYY
Done	<input checked="" type="checkbox"/>
History	<input checked="" type="checkbox"/>
LastName	Savinova
FirstName	Olga
Sex	2
Day	4
Month	April
Year	1964

Figure 5. Recognized form.

The recognition results are displayed in the Form window.

- *Text fields* contain recognized text. Standard keys are used to edit them.
- *Checkmarks*, selected on the image are checked in the recognized form, and selected checkmarks are empty.
- If this or that checkmark was selected in the *radio group*, then in the recognized radio group you will see its number. If no checkmarks were selected, then you get “not selected” in the recognized form. If more than one checkmark were selected, then you get “multiple selection”.
- *Manual input fields* do not contain text, since they are reserved for manual filling from the keyboard.
- The result of *Bar-code* recognition is the recognized code.

Checking recognition results

There is a special mode to check recognition results when the character's image and its recognition result are displayed side by side, which makes the checking

process more efficient and convenient. Each field's contents can be checked in the mode of *context* and/or *group* check.

During the *context* check (you can see the "Field..." title in the window) each character is displayed in its closest context (word or field where it occurs).

During the *group* check (you can see the "Group..." title in the window) each character is displayed together with the alike characters, e.g., all "fives", or all zeros, etc.

Each field included in the context check can be also included to the group check. The group mode is mostly used for digit fields.

The default size of the checked fragment is a word. This means that when you press ENTER the current word is considered checked.

How to check recognition results

1. Select the necessary pages and start checking.
2. When checking the recognition results, press ENTER to confirm that the fragment is correct, check recognition errors, specially mark (press the space key for) the characters which you want to return to later (to check them).
3. Check all the pages containing not checked characters (the pages which are not marked with the check mark in the batch) and all the pages which have the errors in rules (which are marked with a yellow flag in the batch).

To run the check of the text:

1. Click the arrow to the right of the *4-Check* button.
2. From the displayed menu, select *Check recognition results...*
3. You will see the dialog for recognition results checking (for details see below).

Working with the dialog of results checking

Below you can see the dialog which appears when you check recognition results.

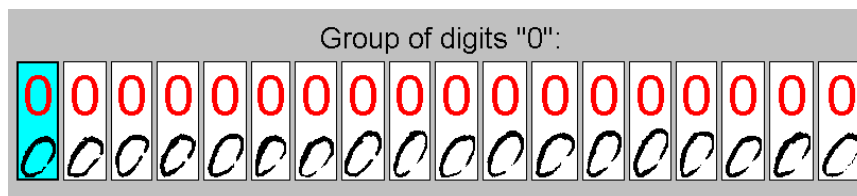


Figure 6. Dialog of results checking during the group check.

Signs in the dialog of results checking

Sign:	Meaning:
Rectangle	Line break
[]	Checkmark
Red character	Indefinitely recognized character
Blue character	Character marked as "to be checked"
Black character	Already checked or accurately recognized character
Yellow-marked character	The character left for further checking (not corrected)
Blue-marked character	Current character (image of the character enclosed in the rectangle)

How to work with the dialog of results checking

To:	Do the following:
Move to a certain character	Press "Right arrow" or "Left arrow".
Correct unrecognized character	Move the entry point to the unrecognized character and click the desired key.
Delete unnecessary characters (including spaces)	Press DEL.
Replace two (or more) characters with one	Select the characters to replace (SHIFT+ "Right/left button") and press the necessary key. The characters will be merged in one *.
Insert a space after the current character	Press INS *.
Insert additional character after the current character	Insert a space (press INS), move to this space and press the necessary key *.
Move the entry point to the beginning or to the end of the fragment	Press HOME or END, respectively.
Confirm a fragment ** (automatically replace all	Press ENTER.


* Available during the context check only.

To:	Do the following:
red and blue characters with black characters)	
Close the dialog without saving changes made after you have pressed ENTER for the last time	Press ESC.

Checking with validation rules

Validation rules are used for automatic check and modification of recognition results. The rules are specified when the template is created. After the page is recognized the rules (if any) are checked automatically.

Errors in rules: signs and their meaning

The page which contains rule errors is marked with a yellow flag (). If there is an error in this or that rule, then it is necessary to check fields included in the rule (whether they have recognition errors) and edit them in compliance with this rule. FineReader allows you to move from one error to another (see below) and edit the contents of the fields included in the rule.

Moving to errors in the rules

For an open page you can see a list of rule errors in a special window under the recognition results of the form.

To move from one rule error to another, use the F8 and SHIFT+F8 keys. This command places the entry point in the beginning of first field of the next or previous rule, respectively.

When the current field loses the focus or when you close the page, the rules are automatically checked. If there are no more errors left in the rules on the page, the error sign (yellow flag) opposite the page will disappear.



Navigating in the recognition results

Checking the rules

To move from one rule error to another, use the F8 and SHIFT+F8 keys. This command places the entry point in the beginning of first field of the next or previous rule, respectively.

****** The size of fragment which is confirmed with the ENTER key, is specified in the form template (in the text block properties). The default size is a word.


Moving to indefinitely recognized words

To check and edit recognition results you can use  or  buttons, to move to the next and previous words, respectively. You can also use hot keys: F4 and SHIFT+F4.

Typical problems when editing

The template was not matched and the page was not recognized, or incorrect template was matched

In this case you should make sure that the template is matched accurately * (see the table) and re-recognize the page.

Reason:	Solution:
Incorrect orientation of the input image. For example, the template has a portrait orientation, and the image is saved with landscape orientation or rotated upside down.	Rotate the image to the necessary position and re-recognize the page. * Buttons:  Error! Filename not specified., Image menu.
Incorrect template is matched or no template is matched with the image.	Select the necessary, check (and edit) the result of template matching and re-recognize the page. * <i>Scan&Read menu> Match template named...</i>
The image is flipped vertically or horizontally.	Flip the image and re-recognize the page. * <i>Image menu.</i>
The image is inverted (the image background is black and the letters are white).	Invert the image and re-recognize the page. * <i>Image menu.</i>

Errors in rules...

Reason:	Solution:
Recognition errors in the rule fields.	Correct all recognition errors in the necessary fields.
Errors in the source document.	Lay the document aside for the



* But if the problem persists, contact your program administrator

Reason:	Solution:
	following check.
Incorrect template matched with the image – it can cause recognition errors and, thereafter, rule errors.	Specify the necessary template and re-recognize the page. <i>Scan&Read menu> Match template named...</i>

For details on rules, see “Validation rules to check recognition results“, p. 51.

The program will not export pages with rule errors.

Large number of recognition errors

Reason:	Solution:
Incorrect orientation of the input image. For example, the template has a portrait orientation, and the image is saved with landscape orientation or rotated upside down.	Rotate the image and re-recognize the page. Buttons:   , <i>Image</i> menu.
Incorrect template matched with the image.	Select the necessary template and re-recognize the page. <i>Scan&Read menu> Match template named...</i>
Low quality of the scanned image.	Set the correct scanning parameters and rescan the image. <i>Tools menu> Scanner parameters...</i>

Export

The program can export only the pages which do not have error flags (neither red nor yellow). If the page is recognized (i.e. has text), then after it is edited the program checks whether the fields' contents corresponds to the rules. The page retains the yellow flag and can't be exported till the validation rules contain at least one error.

To export recognition results to a file:

1. From the *File* menu, select *Export to file...*
2. Name the file where you want to save recognition results.
3. Specify whether to save all recognized pages or only selected. Specify whether the recognized pages are grouped to files according to the templates, or all of them are placed in one and the same file.

For DBF-files only: If there exists a DBF-file, where you have already exported some pages recognized with a certain template you can add other pages to the same DBF-file.

To export recognition results to database:

1. From the *File* menu, select *Export to database...*
2. Check whether the export database is specified correctly.
3. Specify whether to export all recognized pages or only selected.

P a r t 2

Administrator's guide: Tuning the program

Software and hardware requirements

What you will need to work with FineReader 4.0:

- A computer with an Intel 486 processor or higher (Pentium 133 or higher is recommended).
- One of the following operating systems: Windows 95 (Pan-European version is recommended) or Windows NT 4.0 (Service Pack 3).
- 16 Mb of RAM for Windows 95; 32 Mb for Windows NT.
- 30 Mb of free space on the hard disk for minimal configuration and about 80 Mb – for maximal configuration.
- A TWAIN-compatible scanner or analogous device.
- A mouse or other pointing device.
- A CD-drive.
- Color monitor

Installing FineReader

FineReader takes you through installation with onscreen instructions at every step. Before the installation close all anti-virus programs. For best results, do not run any other programs (e.g., MS Outlook, MS Word, etc.) during installation.

1. Insert a hard lock in the parallel port joint.
2. Insert the CD-ROM in the CD-drive.
3. Click *Start* on the Taskbar and choose *Settings/Control Panel*.
4. Double-click *Add/Remove Programs*.
5. Select the *Install/Uninstall* tab and click *Install...*
6. Follow the onscreen instructions.

Note: Items 3-5 are equal to running the install.exe from CD-ROM.


In rare cases problems can arise during installation because of incompatibility of different software components. If the program displays an error message, see the Readme.hlp (you will find it on the CD-ROM) for recommendations helpful in most cases.

Installing and tuning the scanner

Installing the scanner

Important! To connect a scanner correctly, read the user's manual of the scanner. Connecting the scanner, do not forget to install all software bundled (TWAIN driver and/or scanning program).



Create a new batch (click ). For details see "Creating a batch", p. 34.

To start scanning in FineReader, click the *I-Scan* button.

Your scanner should start working immediately, and in 20 seconds you will see a "photograph" of the scanned page on the screen.

If the scanner doesn't start working, you will see the scanner parameters dialog box.

Note: Different scanners have different interfaces. Scanner manager dialog (a built-in TWAIN interface of scanner), is developed by the scanner manufacturer. That is why it is different for different scanners and not always friendly and intuitive.

Now check the scanning parameters (see "Main scanning parameters", p. 28).

How the scanner works with FineReader

The scanner can work with FineReader directly or via a file.


For direct work of FineReader with scanner you need a TWAIN-compatible scanner. TWAIN is a standard protocol supported by the most scanner manufacturers and guarantees a normal work of scanners with FineReader.

To work via file (when images are scanned and saved with the software bundled with the scanner, and after that a batch is assembled from them (see "Assembling a batch from image files")) your scanner doesn't require TWAIN support. Besides, many high-speed scanners do not support TWAIN.

Note: You can view the list of tested TWAIN-compatible scanners in the readme.hlp file (you will find it in the root folder of CD-ROM) or on the web-page of ABBYY (www.abbyy.ru).

Assembling a batch from image files

When working via file you can assemble a batch automatically. To do this, put images in the batch folder. The names of the images to be assembled to the batch must consist of the batch name and the 4-digit number. To update pages in the Batch window choose *Update page list* (View menu) – hot key is F5. The images should be saved in the .TIF format with the same resolution as the batch template image.

You can also open the necessary files from the batch. To do this, open the necessary batch or create a new one. After that click  and select the image files.

If you scanner doesn't work with FineReader

If your scanner doesn't work with FineReader, try the following:

1. Get a new version of the driver from scanner distributor or via the Internet.
Note: If you do not know the address of the manufacturer of your scanner, address to the TWAIN organization committee: www.twain.org. Besides there exist independent TWAIN-drivers manufacturers for the most of popular models, e.g., CFM (www.cfm.de).
2. Go to web-page of ABBYY: <http://www.abbyy.ru>. Perhaps you will find there the new config file for your scanner.
3. Write to technical support service of ABBYY: support@abbyy.ru.
4. If you can't get a new version of the scanner driver, scan the necessary images with the scanning software, save them in the format supported in FineReader (TIFF, BMP, JPEG, PCX, DCX*) and open these images in FineReader (see "Assembling a batch from image files", p. 26).

Setting scanning parameters

Scanning parameters can be set in two ways: from the *Scanner parameters* dialog (*Tools>Scanning parameters...*) or from the TWAIN-driver dialog.

Setting parameters in the TWAIN dialog

You can set scanner parameters from your scanner's dialog. To do this, turn on the *Show TWAIN-driver dialog* (*Tools>Options...>Scanning*) in FineReader. The options of this dialog are described in scanner documentation. For some scanner models the *Show TWAIN-driver dialog...* is switched off by default.

There are scanner models with which you can get good results only when using the TWAIN-driver dialog. When the option is switched off, some TWAIN-driver models may not allow you to set brightness manually and/or choose paper size.

Scanning without TWAIN dialog

When the TWAIN dialog is not displayed, you can set options from FineReader.

The advantages of this working mode are that scanning options are saved for the batch and that you can scan in cycle with a specified pause. Besides, some scanners (e.g., HP PictureScan) allow setting resolution different from 300 dpi only when you switch TWAIN-interface off.

* The list of supported input formats is given in Appendix.

Main scanning parameters

There are **three main parameters**, which influence the image quality, and hence – the quality of recognition. These are *resolution*, *brightness* and *image type* (or *scanning mode*).

Set the correct scanning parameters:

- Resolution – 200 dpi, if you do not need to recognize small (10 points and less) typographic text, otherwise – 300 dpi.
- Brightness (can be also called “threshold” or pictured with the “sun” icon or black-&-white circle, etc.) – set the value so that the background disappears, but the letters are not “broken”. For handprinted texts you usually have to maximally darken the image, so that the letter stroke be not “broken”, but on the other side, not too much, to get rid of the background.
- Image type (scan mode) – should be black-and-white (Line Art, OCR, Text) in most cases. As for gray mode, it is recommended to use it only when the brightness of the image is changing from form to form, or if some blocks on the image have dark (gray or color) background, or if the form background has different brightness.

Attention! To speed up the scanning, narrow the scanning area to the minimum (only check that all necessary fields are included).

After the scanning parameters are set, click the *Scan* button (it can also has some other name, e.g., *Final*), and your scanner will start working.

When the scanning is over you will see the scanned image in the main window of FineReader.

If you have any questions, address to the scanner user's guide.

Checking the scanned image

You can check the properties of the image you have got. To do this, click the **right** mouse button on the image in the *Image* window and from the local menu select *Properties...*

Scanning multiple pages

Working with ADF mechanism

To scan a large number of pages, it is convenient to use the automatic document feeder (ADF).

To make the scanner take pages from the ADF bed, in the *Scanner Parameters* dialog (*Tools>Scanner Parameters...*) select the *Use Automatic Document Feeder*

checkbox. Or, if you work with TWAIN dialog, select the *MultiPage* mode (or alike).

Scanning without ADF

If by this or that reason you can't use ADF, you can scan in the following way: after a page is scanned, the scanner pauses (the length of the pause is specified in options), so that you can insert the next page in the scanner. After that the scanning is automatically resumed.

You only need to specify the pause length. In the *Scanner parameters* dialog (*Tools> Scanner parameters...*) select the *Pause between pages* checkbox. After that specify the length of the pause in seconds. This is additional time added to the pause between the end of one page scanning to the beginning of the next page scanning.

Scanning different types of forms

Color form

If the form has a color background you want to get rid of after scanning, use a filter when scanning. For example, if a form has a red or green background scan with red or green filter, respectively (the filter is specified in the TWAIN interface).

Form with gray or dirty background

If there is a lot of "garbage" on the background or if the background is gray, you have to clean it.

Images which require cleaning should be assembled in one batch. To clean them automatically, turn on the *Clean images when opening* option (*Tools> Options...> Template matching* tab).

But do not make an excessive use of cleaning, since the dots, commas and thin elements of letters can disappear after it, which will inevitably lead to recognition quality decrease. As for handprinted texts, they can be cleaned only if they are not too light.

If the form background is not homogeneous (some parts are "dirty", some parts are light) then you should clean not the whole image, but only separate blocks.

Whether to clean the block is specified in block's parameters when you edit the template (*Properties* dialog> *Block* tab> *Advanced options> Clean block*).

Numbering pages in the batch

You should number pages added from the scanner only if the input documents are numbered or if the form is two-sided. The default number given to the page added from the scanner is the greatest number in the batch plus one. To have an

opportunity to change page numbering select the *Prompt for page number before scanning* option in the *Scanning* dialog (*Tools> Options...*). Then before each scanning operation you will see the *Page number* dialog (Figure 7).

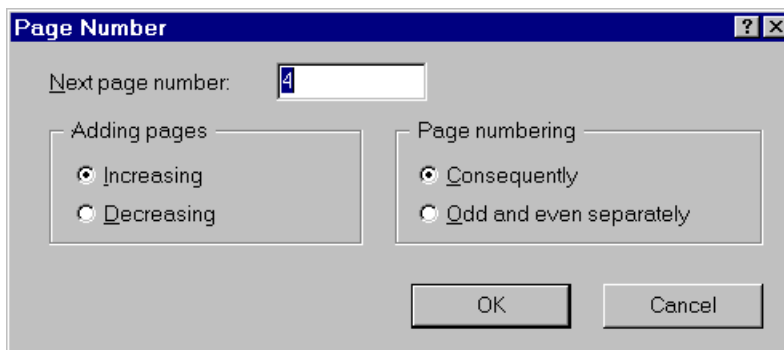


Figure 7. Choose the numbering mode of scanned pages.

In the *Page Number* dialog enter the number of the first scanned page. The pages can be numbered in an ascending or descending order. They can also be numbered consecutively or non-consecutively (only even or only odd) numbers. If, for instance, you use an ADF mechanism to scan a pile of two-sided pages (the pages where the text is printed on both sides) sorted by number, we advise you to scan all pages first from one side, and then from the other side. In this case you should select only odd- or even-numbering. Whether to select ascending or descending order depends on how you place the pile in the ADF bed – starting from smaller or larger numbers.

Machine-readable forms

What does machine-readable form (MF) mean

Machine-readable form (MF) is a regular paper form (questionnaire, blank, etc.) created according to the requirements, specified below.

Like a regular form, MF is meant for gathering structured *primary data* from the person who fills the questionnaire in. The information arranged in *fields* (or requisites), e.g., Last name, First name, Passport series, Date, etc. MFs allow to enter information to databases (DB) automatically with the help of scanners and OCR systems for bulk input of documents.

MF can be filled in: 1) by hand or with a printer/typewriter; or 2) only with a printer/typewriter.

FineReader can input information of the following type:

Types of information	Notes
Handprint characters	Block letters written by hand (Russian, English, German, Ukrainian)
Hand-written characters	Regular handwritten digits or digits written according to a special pattern, as on postal envelopes
Printed text	Typographic text, matrix printer, typewriter (see "Supported languages of recognition", p. 64).
Checkboxes	"Checkmarks", e.g., <input checked="" type="checkbox"/> , <input checked="" type="checkbox"/>
Radio buttons	One of several checkboxes, selected in the group, e.g., <input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> , <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>

Further on you will find requirements for MF filled in by hand and/or with matrix printer/typewriter.

General requirements for MF

1. All MF of one and the same type should be printed from one and the same original (the geometry position of objects should be standardized relative to each other and to the edges of the sheet). The admissible shift of the form elements is 0.15% (half-millimeter for A4).
2. MF must have *reference blocks* which help the OCR system find necessary objects on the image after scanning (the OCR system has to compensate for a skew, for scanning defects, and match the template). For requirements for reference blocks see below.
3. *Descriptive information* (any information and all graphic elements on an empty (not filled-in) form: headings, field titles, black lines, pictures, reference blocks, etc.) should not interfere with subject information. For requirements for descriptive information see below.
4. As for fields containing important information, it is recommended to guarantee an *excess of subject information* to control whether the form is filled in and recognized correctly. The excess of information is achieved with the help of the following fields: Sum in digits – Sum in writing, Control sum, Customer's name – Customer's settlement account, etc. For details see below.

Requirements for reference blocks

For reference blocks you can use special objects – *black squares*, *black lines* or *static text* (descriptive information text).

Different types of reference blocks have different degree of reliability. *Black squares* are the most, *black lines* are less and *static text* is the least reliable. If the form contains black squares, you may not mark out reference blocks of any other type.

Requirements for “black squares”

1. The size of black squares should be from 4x4mm to 7x7mm.
2. Rectangles are not admissible.
3. The forms of one and the same type must have squares of the same size.
4. The optimal number of squares is 5: four squares in the corners of the blank (when you connect the centers of these squares you should get a rectangle) and one square on one side of this rectangle (for automatic detection of image orientation).
5. The distance from the edge of the square to the nearest object (text, line, picture, ...) should be not less than 3mm.

Requirements for descriptive information

As it was already said, *descriptive information* (any information and all graphic elements on an empty form: headings, field names, black lines, pictures, reference blocks, etc.) should not interfere with the subject information (useful information in the form).

The descriptive information (DI) can be of two types: 1) information that will be retained after scanning, 2) information that will disappear after scanning.

“Orange form”

The ideal case is when all DI except reference blocks disappear. To achieve this effect, the DI is printed in light-orange color and the forms are scanned in the Red-Lamp mode (the most high-speed scanners have this mode).

Black-&-white form

If you have black-&-white forms, the *input fields place-holders (IFP)* should disappear after scanning. For this purpose IFPs should: 1) either be light-gray, 2) or consist of small black dots.

Input fields place-holders (IFP)

The fields for subject information should look like the sequence of white rectangles on the gray background (Figure. 8). The rectangles should be not smaller than 4x5mm. The horizontal distance between the rectangles should be not less than 1mm and the distance between lines – not less than 4 mm.

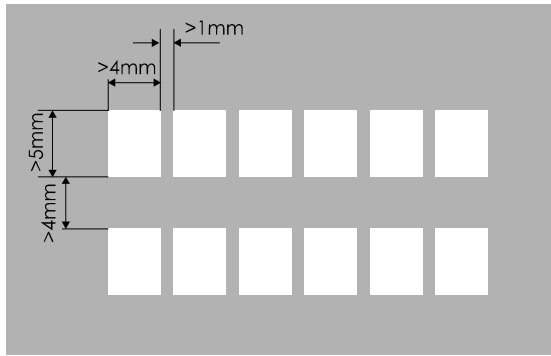



Figure. 8. The scheme of recognized rectangles.

Gray background should: 1) either be very light (5% of black), 2) or consist of separate dots not greater than 0.1mm. The distance between the dots is to be about 1mm.

Requirements for checkmarks and radio buttons

1. The size of checkmarks can vary from 2mm x 2mm to 10mm x 10mm.
2. The distance from the edge of the checkmark to the nearest object (another checkmark, text, black line, picture, ...) should be not less than the size of the checkmark.
3. FineReader can recognize checkmarks of any form: square, circle, rhombus, letter, etc. and marked in any way: checkmark, cross, filled in, encircled, etc.

However, the most reliable shape is a crossed square .

How the customer fills the form in

Forms should be filled in accurately with capital letters with a black ball-point or micro pen. Blue pen is also permissible. The soft-tip pen will lead to lower recognition quality because of filled-in small details. Forms filled in with a pencil or a light color pen are the worst for recognition. That is why it is advisable to type the following words on the form:

“ATTENTION! Please, fill in with BLOCK CAPITAL LETTERS with BLUE or BLACK ball-point or feather pen.”

Auxiliary notes

The form can have areas for auxiliary notes. This field is filled in by the operator. Such fields can include: document number, application date, individual number of

the operator. These marks can be printed with cash-till, with a printer or filled in by hand. They are recognized and exported to database.

Certification

When the number of printed MF copies is rather large, it is advisable to hand the original form over ABBYY/BIT Software for certification. This service is free. The experts will test the original form for machine readability taking into account the experience of the input of many thousand forms and the information about typical mistakes made when the form is being filled in.

When the certification results are positive, you can mark the form blank as:

"Satisfying the requirements of machine readability of FineReader".

General scheme of tuning the system

1. Create a batch for each type of input documents (for details see below, "Creating a batch").
2. For each batch, create and tune templates. The creation of one and more templates is described below, see "Creating template", p. 36.
3. Set scanning, processing and export options for each batch (see "Setting batch options", p. 35).
4. Test the work of the whole system.

Creating a batch

The algorithm of batch creation

1. Click the *New...* button.
2. Name the new batch.
3. Select the image for template (scan paper image or specify the existing image file) *.
4. Specify where to export recognition results (in the beginning you can specify the default value (export to file), and after that edit the template: specify a database or an Automation-object).

* If you have made a mistake when selecting an image for a template, you can change the image (*Template/Change template image...*). But bear in mind that the resolution of the old and the new templates should coincide.

5. Create a template: mark out fields for recognition and not for recognition, specify parameters for the field for recognition (for details see "Marking out fields on the template image", p. 37).
6. Set validation rules (if necessary).
7. Specify additional fields where information about the batch will be placed (if necessary).
8. Save the template.
9. Test the template at least on 20-30 batch images.
10. Edit the template and save it.
11. If you recognize forms of different types in one and the same batch, create additional templates for the same batch (see "Creating template", p. 36).

Setting batch options

Attention! All settings and templates are saved with the batch (except for the settings made in the TWAIN interface). Settings for different scanners are saved independently.

Batch window view

The Batch window can display or not display page parameters. The most convenient view for forms processing is the view with parameters (Figure 9), so that the operator can see the each page's state.

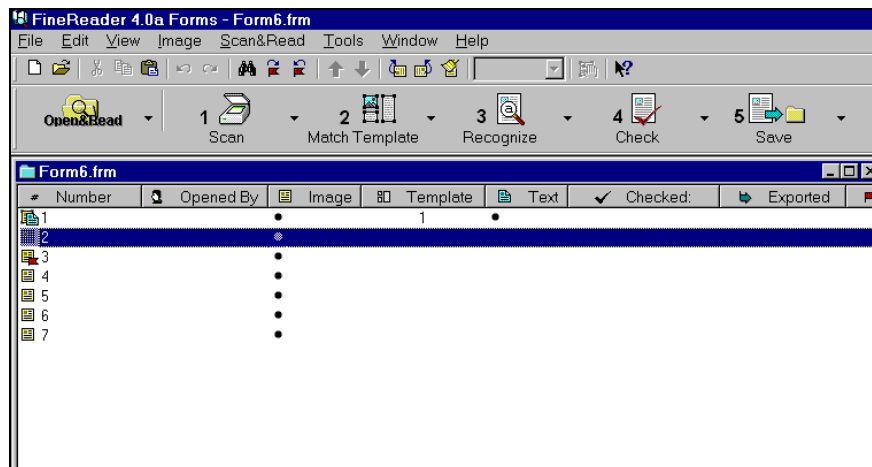


Figure 9. Undocked window. View with parameters.

The Batch window can be docked to the left edge of the Main window or undocked.

To change the Batch view, click the right mouse button in the Batch window and in the local menu select (deselect) the following: *Display page parameters* and/or *Docking view*.

The settings concerning window view, colors, font size and path to the files used in the batch (templates, patterns, languages and rules) are specified in the *General* dialog (*Tools> Options...*).

Scanning

(*Tools> Options...> Scanning*)

- If you are scanning in gray, you can specify that the batch retains only black-and-white images (to save disc space).
- If the input documents are numbered or if the form is two-sided, you may need to number pages. Turn on the *Prompt for page number before scanning* option to display the page numbering dialog before scanning.

Template matching

(*Tools> Options...> Template matching*)

- For maximal automation of image processing we can specify that the orientation of images is detected automatically. That means that when matching the template with the current image, the program will check, whether the image has correct orientation, and if not – rotate the image into the necessary position. This will allow to get rid of “not matching” of template when the image orientation is not correct (e.g., when the image is scanned upside down).
- If all images added to the batch have a dirty background, then you can specify that the images are automatically cleaned when added.

Formatting

(*Tools> Options...> Formatting*)

- Specify the fonts for recognition results representation: a special font for serif and a special font for sans serif font.
- You can specify options for export of recognition results into a file. If you are going to export pictures, specify their format (*Formats...* button).

Creating template

The form template describes position, sizes and attributes of fields on the page. Before the recognition the form template is matched with the current image. The

template is matched with the help of reference blocks (the image shift and the skew up to 10 degrees are compensated for). The template contains information what fields should be recognized.

Algorithm of template creation

1. Mark out reference and text fields, achieve an accurate template matching.
2. Specify languages, dictionaries and other recognition options for each recognized field, test the recognition.
3. Specify the checking options, test the checking procedure.
4. Create additional fields, specify validation rules, test their work.
5. Specify export options, ODBC sources, test the export.

The first template of the batch is created together with the batch itself. After the batch and the first template are created, you can make additional templates to be used in the same batch (for details see "Creating template", p. 36).

Preliminary editing of the form template image

After you have selected the image for the template and it is loaded in the *Template editor* window, you can edit it if necessary: e.g., rotate, invert, clean it, etc. To do this, select the corresponding items from the *Image* menu.

Marking out fields on the template image

Field types on the template image

The form template includes fields of three types: reference (not for recognition), for recognition and auxiliary (additional and control) fields.

Reference blocks are used to match template with the image, *blocks for recognition* are recognized. *Additional fields* are used to export information about the page to database (e.g., when it was created, recognized, etc.). *Control fields* are recommended for accurate matching of template when the batch contains forms of different kinds.

Reference blocks

Reference blocks* are used to match template with the image. These are *black squares*, *black lines* and *static text*. They enclose the parts of the image which occur on each page and do not contain information to input in the database.

* Saying reference blocks we mostly mean black squares, since they guarantee maximal accuracy when matching template.

- *Black squares* are usually positioned in the corners of the blank and guarantee the utmost accuracy of template matching.
- *Black lines* – vertical or horizontal lines on the form.
- *Static text (field)* – descriptive text, which occurs on each blank and can be used to match template with the image. When the batch contains forms of different kinds, static fields can be used for accurate template choice.

Blocks for recognition

Blocks for recognition can be of the following types: text fields, checkmarks, radio groups and bar-code.

- *Text fields* – fields containing subject information after the blank is filled in.
- *Checkmarks* – squares which can be selected by “cross” or “checkmark”.
- *Radio groups* – two or more checkmarks, from which only one checkmark should be selected (e.g., yes, no, don't know).
- *Bar-code* – a code to denote a country, an enterprise, etc. Its recognition result is digits and letters.

Selecting fields on the form template

Important! To create a template, it is recommended to take an empty (not filled in) blank, so that static fields can be marked out automatically and fields for recognition can be “drawn” by hand.

1. First of all mark out reference blocks automatically. To do this, from the *Template* menu select *Automatically mark out blocks of certain type...* In the displayed dialog (Figure 10) specify what blocks you would like to mark out on the image.
-

Attention! If the form contains black squares, then it will be enough to mark them as reference blocks. If there is a sufficient number of black lines (2-3 vertical lines and 3-4 horizontal lines), then you may not select static text. Do not mark out the objects which may exist not on all images.

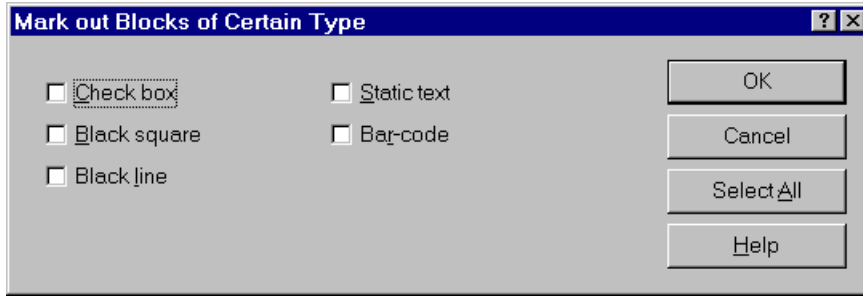


Figure 10. Dialog for block marking on the template image.

2. Edit the marked blocks, “draw” text blocks and, if necessary, unite selected checkmarks into groups.

How to “draw” a block:

Position the mouse pointer in the upper left corner of the block you want to “draw” on the form template. Drag diagonally until the frame encloses the necessary part of the image. Release the mouse button. The selected part of the image will be enclosed in the green frame (by default).

How to unite checkmarks into groups:

The selection of group is similar to text field marking. First of all check whether the checkmarks you want to unite are marked out and have the *Checkmark* type. Make certain that these checkmarks are not exported (the *Exported block* checkbox is not selected) After that draw a block which will enclose all checkmarks you want to unite in one group (see above – “How to draw a block”).

3. If you want to export additional information about the page to database, create an additional field in the form template (see “Additional fields“, p. 42).
4. After the fields are marked out, you should specify parameters for fields to be recognized (for details see “Text field parameters“, p. 39).

Deleting fields

To delete a field, click it with the mouse and press DEL. You can also delete all blocks of a certain type (*Blocks> Delete blocks of the type...*).

Text field parameters

If you look at the form you can see that a number of blocks for recognition have alike parameters, and other blocks, vice versa, have individual parameters – in contrast to the most blocks in the template. For example, the most blocks on the form must have Russian language of recognition, the *Auto* type of text, etc., but there are such blocks which require another text language, e.g., digits or Russian-English, etc.

First of all specify the parameters appropriate for most of text blocks on your form. Set these parameters as default for the form template (*Properties* dialog, *Default settings* tab). To recognize other blocks, you need to specify individual parameters, different from *Default* parameters.

Default parameters

First of all specify default parameters, i.e. parameters appropriate for most of blocks (*Properties* dialog, *Default settings* tab). These are parameters which will be automatically set when a text block is “drawn” on the form.

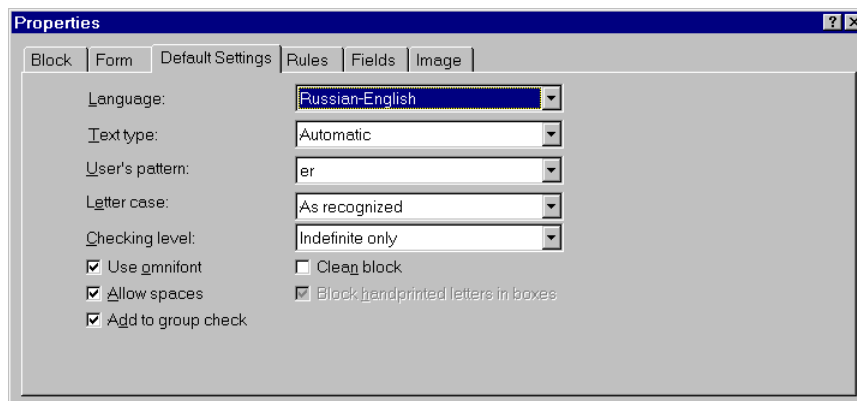


Figure 11. Default parameters for text blocks.

Individual parameters for text field

If the parameters of this or that block are different from the default settings, deselect the *Default settings* on the *Block* tab of the *Properties* dialog and specify the necessary values.

Options set for text field can be divided into three groups: *General*, *Advanced* and *Checking* options. For details on these options see below “General, Advanced and Checking options”.

To choose between General, Advanced and Checking options:

1. In the *Template editor* dialog click the field you are interested in.
2. On the *Block* of the *Properties* dialog select the necessary value from the *Options* list.

General, Advanced and Checking options

General options

The main parameters you should set for each text field are language (i.e. words of which language can occur in this field), text type (auto (for typographic texts), matrix printer, typewriter, handprinted) and database field to export recognition results. If the field is filled in with handprinted letters of a special kind – each letter is written in a special square (for details see “Orange form”, p. 32), then you should also select the *Block handprinted letters in boxes* checkmark.

Advanced options

Advanced parameters include user's pattern (for accurate recognition of fields of bad printing quality). By default, together with the user's pattern the omnifont pattern is used*.

To increase recognition accuracy of fields containing only one word, deselect the *Allow spaces* checkbox. You can also specify letter case, to format contents of the recognized field. For each field you should specify a level of error marking (the default level is standard, when the characters with a very high degree of uncertainty are marked with color).

When forms of different types are recognized in one and the same batch, it is advised to create fields with fixed values, according to which the program would find the necessary template (*Field value*).

If the block has a dirty background on all forms of the given template, you can clean it.

If you recognize “flexible” forms (i.e. the forms whose fields can be shifted relative to each other) you can select the *Detect field position* checkbox. Thus, when the program matches the template according to reference blocks, it will “look for” this field till it finds it in the image.

Checking options

Parameters for recognition results checking.

For each field you can install a checking level (whether to check this field's characters or not, and if yes then what characters: only indefinitely recognized or all characters). For each field you should install the size of the checked fragment (the fragment which is confirmed with the ENTER key): it can be a character, a word, a field.

* It should be switched off only when recognizing letters of unusual shape (e.g., Armenian or Georgian fonts).

For each field you can specify whether to include it in the group check (the check when the character is displayed in the group with alike characters). For details see "Checking modes", p. 50.

Control fields

Control fields are used for accurate template matching. This field specifies the value which occurs in all forms recognized with the given template. When the program tries to match the template with the current image, it compares the recognized value of this field with the corresponding field's value specified in the template.

For *control fields* you can use *text fields*, *bar-codes* and *static text*.

The control field value is specified when you edit the form template.

To specify the value of the text field:

1. In the *Template editor* dialog activate the text field, the contents of which will be the same on all pages recognized with this template.
2. On the *Block* tab of the *Properties* dialog select the *Advanced* value from the *Options* list.
3. In the *Field value* box enter the text which will repeat on all images recognized with the current template.

To specify the value of the static field:

1. In the *Template editor* dialog activate the text static field, the contents of which will be recognized and will be the same on all pages recognized with the current template.
2. On the *Block* tab of the *Properties* dialog enter the value of the static field.
3. After you have specified the *Static field* value, you can be modify its parameters.

To specify the value of the Bar-code:

1. In the *Template editor* dialog activate the *Bar-code* field.
2. On the *Block* tab of the *Properties* dialog enter the value of the field.

Additional fields

Additional fields are the fields where the information about the page is specified. This can be information from the Batch window, e.g., batch number, template name, error number, etc.

To create an additional field on the form template:

- *Tools> Templates...> Properties> Fields tab> Add...* button.

You can also “draw” an additional field on the form image (as a usual block) and make it an *Additional* block.

Manual input fields

The block which has the *Manual input field* type, is not recognized. These are blocks to input information which is written with handprinted letters or not contained on the form.

Such information is input from the keyboard, and the connection with the image is saved for these fields, which noticeably simplifies the input of fields, written in handwriting.

Setting export options

Export to database

Database tuning

1. In the *Properties* dialog (*Tools> Template, Properties...* button) select the *Form* tab (Figure 12, p. 40).
2. In the right part of the dialog there is the *Database connection* field. To change its contents, click *Change...*

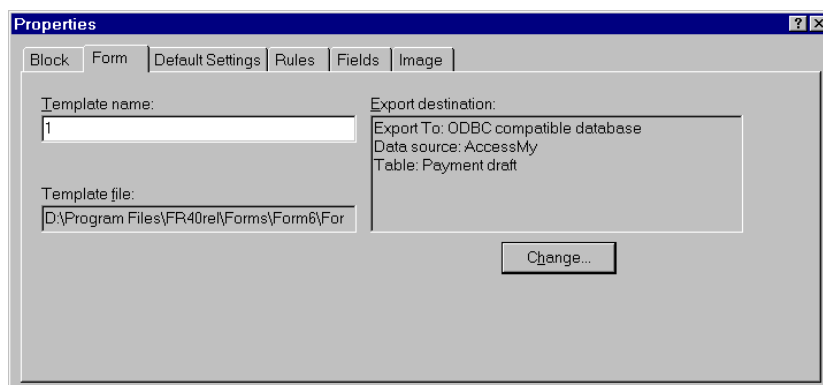


Figure 12. “Form” tab of the “Properties” dialog.

3. You will see the *Export mode* dialog (Figure 12). Set the switch to the *ODBC-compatible database* position and click *Next*.
4. Specify the database source and the table. If there is no table yet, enter the name for a new table, which will be automatically created during export.

Specifying database field for a certain form field

If the export database already exists, select the corresponding database field from the *Database field* list.

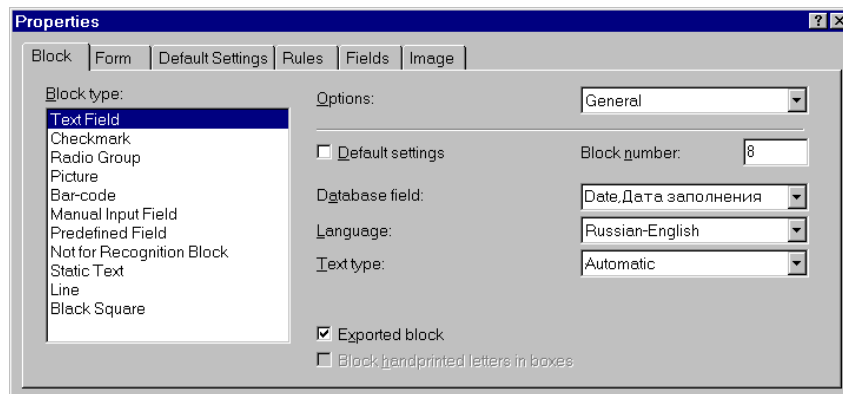


Figure 13. "Properties" dialog. "Block" tab.

If the database is created during export, the names of the future database fields are input manually. The name of block in the template can consist of two parts. For example, "Date, Дата заполнения". In this case "Date" is the name of the field in the database, where the contents of the field will be exported, and "Дата заполнения" is the internal name of the field in FineReader. This mode of block naming is convenient if the database where you export recognition results has some restrictions on the naming conventions of the fields (e.g., only Latin letters and short names).

Export to file

The recognition results can be exported to a file of one of the following formats: TXT (DOS and Windows), CSV (DOS and Windows), DBF, XLS. To do this, during the template creation specify the following export type – *DBF, XLS, etc.* files. (Figure 14).

Attention! Export in DBF is available only if the fields of the form do not contain spaces and are no longer than 10 characters.

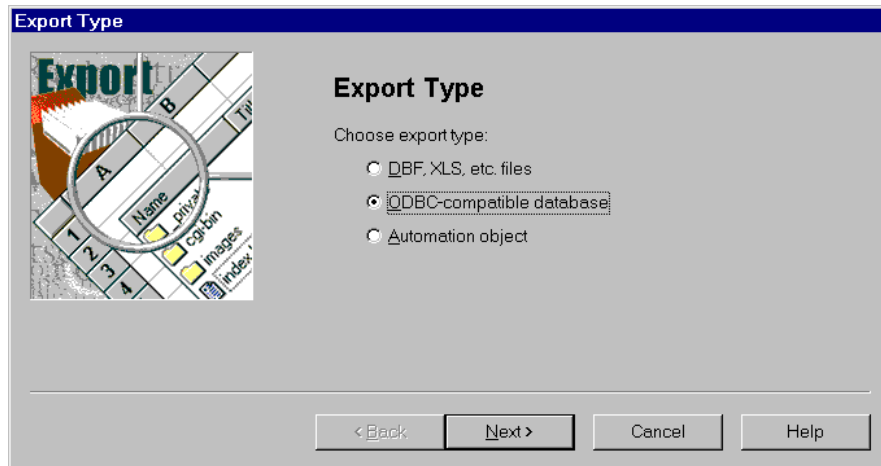


Figure 14. “Export Type” dialog.

Export via Automation

For details on export via Automation see FineReader Automation API guide.

What you get after export

Text fields are presented with their recognized values.

Checkmarks are presented with true and false values (checked/not checked), and radio groups – with the number of the checkmark selected in the group, –1 (if no checkmarks are selected) and –2 (if several checkmarks are selected).

When you export recognition results to a file, pictures are not saved. When you export recognition results to a database, pictures are saved in a binary format in a special field. Picture formats are specified on the *ODBC* tab of the *Formats* dialog (*Tools> Options...> Formatting> Formats...*) (Figure 15, p. 46).

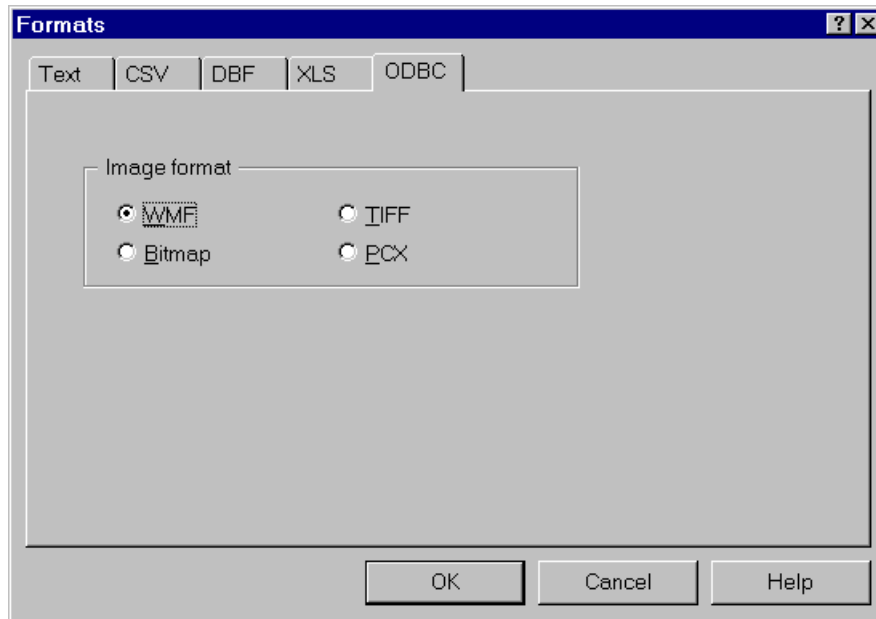


Figure 15. Dialog to specify picture format to export them to a database.

Multiple templates

One batch can contain forms recognized with different templates. The batch stores different templates for each type of batch forms. The necessary template is selected automatically, which is why you do not need to sort forms before recognition.

The first template is created together with the batch. The rest templates are created with a special command (*Tools> Templates...> New...*).

Attention! If you create several templates for a batch, the resolution of these templates should coincide with the resolution of the first template.

For better distinguishing of templates

If the batch contains more than one template, check that reference blocks on the template are quite different. Different templates should have different number of reference blocks; reference blocks should have different positions, if possible (for

details see “Reference blocks“, p. 37). This will make the program choose a correct template for a certain image.

For accurate template matching you can also use control fields (see “Control fields“, p. 42).

User's languages of recognition

FineReader allows you to create and attach user's languages of recognition. These can be single languages or groups of (two or more) languages. Attaching user's languages can considerably increase recognition quality of forms.

When to create a new language

There are three main situations, when it is necessary to create a user's language or a group of languages:

- Attaching a special dictionary from the text file or from regular expression -- for recognition of certain form fields.
- To recognize texts in the languages not supported in the system.
- To recognize fields containing words of several languages.

You can attach a dictionary to the user's languages (see below).

The sample of new language creation for recognition of Old Russian texts is given in User's Guide of FineReader 4.0 Pro.

Creating user's languages and language groups

You can create a new language or language group from the *Language editor* dialog (*Tools>Language editor...*).

What should be specified for user's language

You should specify the following for user's language:

1. Name.
2. Set of available characters (for details see below).
3. Dictionary (if necessary).
4. If you have selected the *User* code page for a new language, specify the font to display the recognition results (*Tools> Options...> Formatting* tab).

What is specified for the user's group of languages

1. Name.
2. Languages of the group.
3. Set of available characters.

Characters for the user's language or user's group of languages

When creating a new language, specify the code page, and then select the characters which will occur in the language.

Digits

Besides digits, the *Digits* language also contains the following special characters: { } " # \$ % () * + , - . / : ; = X [] x ° © № § £ ¥.

Prohibited characters for a user's language

If you attach digits to a new language, then together with digits (1, 2, etc.) a set of special characters is automatically attached (see above).

That is why if the texts you will recognize with a new language do not have some digits or special characters, add them to the *Prohibited* list.

Prohibited characters for a group of languages

Like a new language, the group of languages may have prohibited characters, digits and special characters attached together with digits (see above). Besides these characters you may need to exclude some characters which are contained in one or more group languages, but should not be included in the group.

Ignored characters for a user's language or a group of languages

Ignored characters are characters which occur inside words, e.g., bullets. These characters are marked as ignored – so that the words where these characters occur are not considered incorrect.

Prefixes and suffixes

Prefixes are non-letter characters which can occur in the beginning of words.

Suffixes are non-letter characters which can occur in the end of words.

If you add any characters to the prefixes or suffixes list, the word will be considered correct even if this character is “stuck” to it.

Word separators

Characters which can separate words (instead of spaces), e.g., /, \, dash, etc.

Attaching a dictionary to the user's language

Attaching a dictionary to the user's language can considerably increase the recognition quality. For example, the dictionary attached to the Surname field can contain surnames which are available in this field. For each language you can specify whether it can contain words not only from the dictionary.

You can use a text file (in Windows encoding) as a dictionary. The words in the file you use for a dictionary should be separated with spaces or other characters not included in the alphabet.

You can also attach a dictionary of existing user's language.

Attaching a dictionary of existing user's language (with regular expression)

Specify the expression of the @ (Lang) type. @ (Lang) denotes a dictionary of the Lang language.

Requirements for the Lang language, used in this regular expression:

- A dictionary is attached to it.
- Only dictionary words are available.
- The case of dictionary words is the same as their case in the dictionary (see below).

Specifying the case for dictionary words

In the *Word case* dialog (Figure 16) you need to specify the case which is considered correct for the words of the attached dictionary.

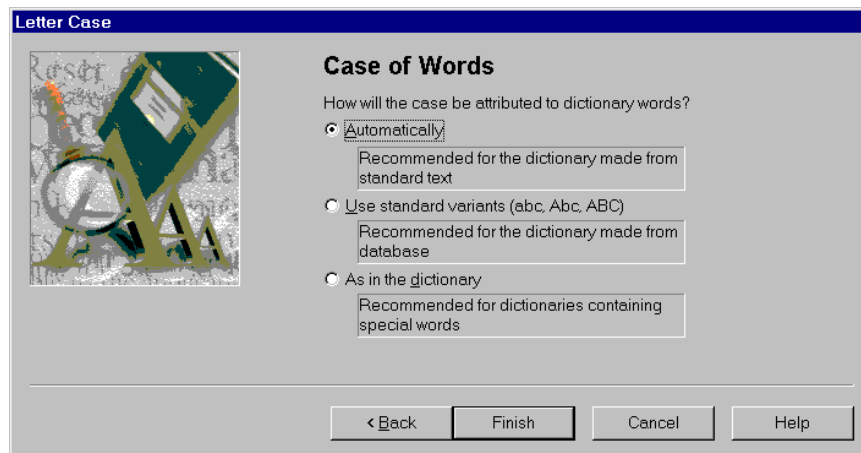


Figure 16. "Letter case" dialog.

The case of dictionary words can be defined in the following way:

1. Automatically, as in a regular dictionary:

If the word in the dictionary looks like following:	Then the following variants of the word will be considered correct:
abc	abc, Abc, ABC
Abc	Abc, ABC
ABC	ABC

If the word in the dictionary looks like following:	Then the following variants of the word will be considered correct:
ABc	aBc, ABC

2. Standard variants:

Dictionary words will be considered correct, if they occur in the text in their dictionary case or with one of the standard cases (small letters, capital letters, with the first capital letter).

3. As in the dictionary:

The dictionary word is considered correct in the text if it has *the same case as it has in the dictionary*.

Checking recognition results

How the check is carried out

There is a special mode of check when recognition results can be viewed and edited quickly and efficiently. During the check the operator can see the image of the checked character and the result of its recognition side by side.

Checking modes

The check can be carried out in two modes: context and group.

Context check is convenient for fields containing words: each character is displayed in its closest context (word or field where it occurs).

Group check is usually used for digit fields: each character is displayed together with the alike characters, e.g., all “fives”, all zeros, etc., which can speed up the check considerably.

A group check is run for checkmarks and radio groups.

Including handprint fields in the check

For each field of the form you can specify whether to include it in the check or not. And if yes, whether to check *all* character or only *indefinitely recognized*. It is advised to specify the checking level used for most of blocks in the defaults tab for text blocks (*Properties* dialog>*Default settings*).


If necessary, you can set an individual (different from default) checking level for a block. To do this, on the *Block* tab of the *Properties* dialog deselect the *Default settings* checkbox and select the value from the corresponding list.

The size of checked fragment is the fragment which is confirmed (marked as correct) when you press the ENTER key. All characters of this fragment are confirmed (considered checked). The fragment can be a character, a word or a

field (*Block* tab of the *Properties* dialog, *Checking* options, *Fragment for checking* list). The default size of the checked fragment is a word.

Each text field can be included in the group check (see above).

Validation rules to check recognition results

Validation rules are used for automatic check and modification of recognition results. The rules are specified in the form template. After the form fields are edited and recognized, it is checked whether they conform to the validation rules. The page with rule errors is marked with yellow flag () in the Rules column.

Validation rules list

- Automation-check
- Replacing values from the list
- Replacing characters from the list
- Normalization and check of date
- Normalization of price
- Fields merge
- Check against database
- Expression
- Sum checking
- Sum in digits – sum in writing
- Regular expression

How to create a rule

Validation rules are specified when you create or edit the form template (*Tools> Templates...> Properties> Rules* tab).

1. In the *Rules* tab of the *Properties* dialog click *New...*
2. Select the rule to create and click *OK*.
3. Name the rule.
4. For the *Check against database* rule you can select the *Reliable check* checkbox. In this case if the rule (and all the rules which include the fields of this rule) has no errors, then the fields will be marked with gray as accurately recognized and not requiring check.
5. Click the *Next* button and specify the necessary rule options. For details about each rule see below.

Automation-check

Automation-check rule allows the user to set his/her own checking options for the form fields, display error messages, if the check was a failure, and modify field values.

For details about Automation-check see FineReader Automation API guide.

Replacing values from the list

This rule allows you to find the necessary fragment in the specified field of the form and replace it with the other fragment.

When creating the rule specify the following:

1. The field where to search and the field where to write the replacement down.
2. What to find and what to replace with.

The field where you search the necessary value, and the field where the replacement is placed can coincide or not. If the fields do not coincide, then the contents of the field where the information is sought is copied to the replacement field, where the replacement is made.

The screenshot shows a 'Validation Rule' dialog box with a blue title bar. It contains two dropdown menus: 'Find where:' and 'Write to where:', both set to 'Field1'. Below these is a table with three columns: 'Find What', 'Replace With', and 'Search type'. The table contains one row with the values 'Ltd', 'Inc', and 'Exact'. To the right of the table are three buttons: 'Add', 'Edit...', and 'Remove'. At the bottom of the dialog are four buttons: '< Back', 'Finish', 'Cancel', and 'Help'.

Find What	Replace With	Search type
Ltd	Inc	Exact

Figure 17. Rule: Replacing values from the list.

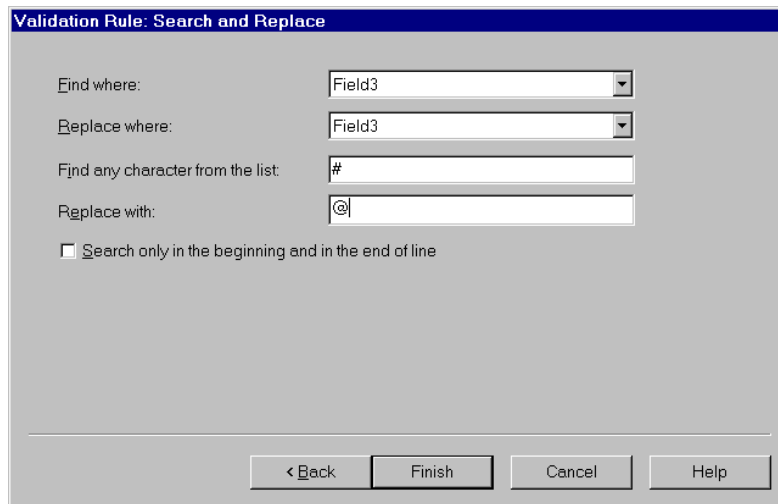
One rule can contain several pairs “fragment to find – fragment to replace with” (Figure 17, p. 52). To add a new pair, click *Add...* You will see the *Search and replace* dialog (Figure 18).



Figure 18. “Search and replace” dialog.

When specifying a pair for replacement, indicate whether to search (and replace) the whole string, and select the search type: exact (exact coincidence of letters and their case), ignoring the case, and fuzzy (the number of available errors is specified).

Replacing characters from the list



The dialog box titled "Validation Rule: Search and Replace" contains the following fields and controls:

- Find where:** A dropdown menu with "Field3" selected.
- Replace where:** A dropdown menu with "Field3" selected.
- Find any character from the list:** A text input field containing "#".
- Replace with:** A text input field containing "@".
- ☐ **Search only in the beginning and in the end of line**
- Buttons at the bottom: "< Back", "Finish", "Cancel", and "Help".

Figure 19. Rule: Replacing characters from the list.

The rule allows to find specified characters in this or that form field and replace each of them with a specified character (or a fragment).

Note that each character specified in the *Find any of the characters* box (Figure 19, p. 54) is replaced with the contents of the *Replace with* box. The characters are not separated by spaces or other characters.

When creating the rule specify the following:

1. The field where to search and the field where to place the replacement.
2. Characters to be replaced.
3. Characters which will replace the sought characters.

The field where you search the necessary value, and the field where the replacement is placed can coincide or not. If the fields do not coincide, then the contents of the field where the information is sought is copied to the replacement field, where the replacement is made.

This rule can be used to delete spaces and different “garbage” (brackets, slashes, etc.), which can appear in the end and beginning of the word. To make the program do it, check *Search only in the beginning and in the end of line*.

Normalization and check of date

Attention! The rule works for Russian only!

This rule can consist of two parts. The first is normalization of the recognized date (to make it look like, e.g., 01.01.99, or 1 янв 1999, etc.). The second part is the check whether the recognized date matches the specified time span. The second part is optional, i.e. you can only normalize the date and not check whether it matches the specified time span.

For date normalization specify the following:

1. The field where to search the date and the field where the modified date will be placed (the fields can coincide).
2. Month and date format.

When specifying the time span

You can set two limits: *not earlier than* and *not later than*. You can count starting from the current date or from any fixed date, e.g., two days (months, years) before the current date, 1 January 1997, etc.

To specify a fixed date, select the corresponding checkbox: for the lower and the upper limits respectively (Figure 20, p. 55); then specify the date.

Figure 20. Checking whether the date matches the specified time span.

Normalization of price

Attention! The rule works for Russian only!

This rule automatically normalizes the price (e.g., 2.50 or 2,50).

When creating the rule specify the following:

1. The field where the recognized price is located, and the field where the normalized price is written down (the fields can coincide).
2. The character which will separate rubles and kopecks in the field.

The field containing the recognized price allows the following words besides the digits “рубль (рублей, рубля, руб.)” and “копейка (копеек, копейки, коп.)”.

If there is only one character after separator (e.g., 27.3), a zero is inserted before this character (27.03).

Merging the fields

This rule allows to merge in a certain form field the recognition results of several other fields of the same form.

For example, if the database where you want to export recognition results, contains the “Last name, first name” field, but the form has two separate fields corresponding to it (First name and Last name), then the information from these fields can be merged in one field of the form and after that the contents of the merging field will be exported to a database.

When creating the rule specify the following:

1. The field whose contents will be merged *.
2. The merging field.
3. Separator (character separating the contents of the merged fields) ** (Figure 21).

* The order in which the contents of the merged fields is placed in the merging field depends on the order they are listed in the rule (from top to bottom) (Figure 21, p. 57).

** If the character is not specified, the contents of the merged fields is not separated and is thus stuck together. If you do not need any special separator, enter a space.

Figure 21. Rule: merging of form fields.

Attention! If the export database is not created yet and you are planning to create it automatically during the export, create an additional field of the *Empty field* type (for details see “Additional fields”, p. 42).

Checking against database

This rule compares the contents of the specified form field with the corresponding field of the database. If the check was a success, the rule is fulfilled. Otherwise there is a rule error.

When creating a rule, specify the following:

1. Database with which to compare.
2. A table from this database.
3. “Form field – database field” pair.
4. Comparison type (see below).

Important! Each *Checking against database* rule should have a field with exact comparison. This will allow the program to find the necessary entry in the database and compare all the rest fields of the given rule with it.

Types of comparison against database:

- Exact comparison:
The content of the form field and of the database field should coincide.
- Exact comparison with suggestions:
If the content of the form field is similar but not exactly the same as the content of the database field, then the program will suggest a correct variant for a form field (the necessary database entry will be found according to the field with exact comparison).
- Database comparison and substitution:
The form field contents is replaced with the contents of the corresponding database field. If there are several variants, the program uses the first in the list.
- Fuzzy comparison:
If, e.g., in the form field you can see “ТОО Гапарж”, and the corresponding database field contains “Гапарж”, then this form field is considered the same as the database field. Fuzzy comparison can ignore such words as, e.g., “город”, “РФ”, “АО”, “ООО”, “ЗАО”, etc.
- Fuzzy comparison with replacement:
If the form field contents doesn't coincide exactly with the contents of the database field, but looks alike (large percent of words are the same), then the contents of the found database field is automatically written over the form field's contents.

Check with expression

This rule allows you to write an expression in a special language. The expression will check (and modify) the contents of recognized fields.

For details about this see FineReader Automation API guide.

Sum checking

This rule adds up digit values of specified form fields and compares the sum with a certain value. This value can be either a certain constant or the value of the specified form field.

When creating a rule, specify the following:

1. Fields which will be added up.
2. Whether to compare the sum with the constant or with the value in the specified field.
3. If the calculated sum is compared with a certain digit value, specify this value. If it is compared with the value of a certain form field, specify the field.

The sum can't be greater than 2147483647 or less than -2147483648.

Attention! Before using this rule, it is advised to run *Sum normalization* rule for the summed fields (see above).

Sum in digits – sum in writing

Attention! The rule works for Russian only!

When creating a rule, specify the following:

1. The field containing the necessary sum in writing (e.g., двадцать пять рублей семьдесят девять копеек).
2. The field where the same sum is in digits. Besides the digits, this field can also contain words like “рубль (рублей, рубля, руб.)” and “копейка (копеек, копейки, коп.)”. The digits, denoting rubles and kopecks can be separated with a dot or a comma.

If the sum in digits is recognized accurately and the sum in writing is recognized not very well, the program suggests a variant of the sum in writing: to the right of the *Sum in writing* field there appears an arrow clicking which you will get a local menu with replacement suggestion.

Attention!

1. The rule works for rubles and kopecks only.
 2. Before using this rule for the field containing the sum in digits, run *Sum normalization* for this field.
-

Regular expression

This rule compares the contents of the form field with the regular expression. In the regular expression we specify which characters are allowed in this or that position of the field.

When creating a rule, specify the following:

1. The field to be checked with regular expression.
2. Regular expression text.
3. Text of error message.

To create a regular expression, click the button to the right of the regular expression field (Figure 22, p. 60).

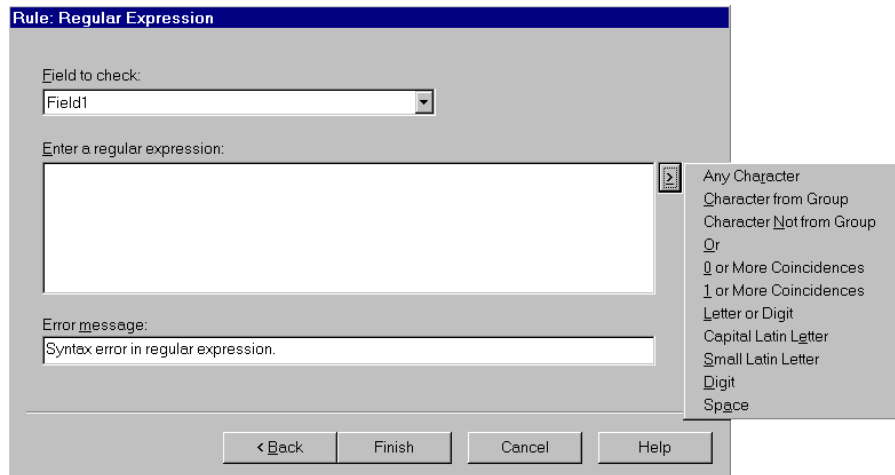


Figure 22. Dialog to create regular expression.

Alphabet used in regular expressions

Name in the list	Character	Example
Any Character	.	s.t allows "sat" and "set" and so on.
Character from Group	[]	w[io]n allows "win" and "won".
Character Not from Group	[^]	m[^a]st allows "mist" and "most," but not "mast." t[^a-m]ck finds "tock" and "tuck," but not "tack" or "tick."
Or		re(e a)d - allows "reed" and "read"
0 or More Coincidences	*	10* - allows 1, 10, 100, 1000 and so on.
1 or More Coincidences	+	10+ - allows 10, 100, 1000 and so on.
Letter or Digit	[0-9a-zA-Za-яА-Я]	[0-9a-zA-Za-яА-Я] - allows any single character; [0-9a-zA-Za-яА-Я]+ - allows any word
Capital Latin Letter	[A-Z]	.
Small Latin Letter	[a-z]	.

Name in the list	Character	Example
Capital Cyrillic Letter	[А-Я]	
Small Cyrillic Letter	[а-я]	
Digit	[0-9]	
Space	\s	

Training a pattern

What does to teach a program mean

Pattern is trained to improve the recognition of forms of poor printing quality.

To train a program (or to train a pattern) means to set up in the OCR memory a pairing “the symbol image – the symbol name”.

Trained and edited pattern is attached to the necessary form fields (when editing the template) and is used during recognition.

For details about training see the user's guide for FineReader 4.0 Pro (FR_Rus.pdf in the Guide folder on the CD-ROM).

Attention! You can train the program only if the text type is not Handprinted.

How the system is trained

This is the set of actions you need to do in order to create, attach and train the pattern for forms recognition.

1. Create (name) the pattern (*Tools> Template editor> New...* button).
2. Select the *Read&Learn* checkbox (*Tools> Options...> Recognition*).
3. Edit the form template: set the pattern for *the fields to be recognized with this pattern*. If this pattern will be used for recognition of the majority of fields, specify it in the *Default settings*, and as for the fields which do not use it, set individual parameters for them (for details see “Individual parameters for text field“, p. 40).
4. Run recognition (*3-Recognize* button) and train the pattern (for details on pattern training see User's guide for FineReader 4.0 Pro (FR_Rus.pdf in the Guide folder on the CD-ROM)).
5. Edit the pattern (*Tools> Pattern editor> Edit...* button).
6. Turn Read&Learn mode off (*Tools> Options...> Recognition> Read&Learn*).
7. Now the pattern is ready and can be used for batch forms recognition.

Batch processing in the network

You can process one batch on several computers **simultaneously**. This speeds up the input of large amounts of documents considerably. Each computer involved in the processing of the batch must have a separate copy of the program installed.

Input of text to the computer involves scanning, page analysis, recognition, editing and export. When processing a single batch simultaneously on several computers, you can spread the task so that one operation or combination of operations can be performed on each computer. For example, one computer, equipped with a scanner and ADF mechanism, can be used for scanning, while another computer (or several other computers) recognizes the scanned pages and/or edits and exports recognition results.

If batch recognition is run on several computers simultaneously, they divide the work between themselves automatically: as long as there are unrecognized pages in the batch, the computers will not stand idle.

How to work with a batch in the network

1. Create a new batch (unless it already exists) and save it in a folder that is read- and write-accessible by all computers involved in the document processing.
2. Create a form template: mark out blocks, set blocks' parameters, create validation rules, specify export parameters, etc. (for details on template creation see "Selecting fields on the form template", p. 38).
3. Check the template (recognize 20-30 images with it).
4. After the template is ready, close the batch to save its options.
5. Run FineReader and open the batch on the computers that will work with it.
6. Start scanning on the computer which is equipped with a scanner and ADF mechanism.
7. Start recognition in the *Read&Wait* mode on the computers meant for recognition (*Scan&Read>Recognize All Unrecognized Pages and Wait*).
8. Once some pages have been recognized, you can start editing them on any free computers, or on the computers where background recognition is running, without waiting for the remaining pages.
9. Recognized pages (all or the selected) can be exported to a file or sent to a database.

What is FineReader Automation API

FineReader Automation API allows to control the work of the FineReader 4.0 OCR system from your program. FineReader can be integrated with existing

document processing programs or used as a base to develop a special application to input and process documents.

The program is controlled with the Automation interface. FineReader can be controlled both using such languages as Visual Basic or C/C++. The interface contains a large number of functions, which allow to control practically all aspects of FineReader's work. FineReader Automation API includes detailed description of the interface, reference-file and samples of API use in Visual Basic and C++.

FineReader Automation API also describes how to export via Automation and write your own checking rule.

To buy FineReader Automation API, address to ABBYY, ph.: +7 095 263-6658, 263-6659.

P a r t 3

Appendices

Supported languages of recognition

Handprint recognition

- English
- German
- Russian
- Ukrainian

Printed recognition

Languages with dictionary support

- Bulgarian
- Czech
- Danish
- Dutch
- English
- Estonian
- Finnish
- French
- German
- Greek
- Italian
- Norwegian
- Polish
- Portuguese
- Russian
- Spanish
- Swedish
- Tartar (in a special delivery)

- Ukrainian

Languages without dictionary support

- Afrikaans
- Albanian
- Basque
- Byelorussian
- Catalan
- Croatian
- Hungarian
- Icelandic
- Indonesian
- Irish
- Lettish
- Lithuanian
- Macedonian
- Romanian
- Serbian (Cyrillic)
- Slovak
- Swahili
- Tagalog
- Turkish

Input image formats

Type:	Color:
BMP	b/w, gray, color
PCX, DCX	b/w, gray, color
JPEG	gray, color
TIFF	b/w – uncompressed, CCITT3, CCITT3FAX, CCITT4, packbits; gray – uncompressed, Packbits, JPEG; color – uncompressed, Packbits, JPEG

Hot keys

	To:	Press:
File menu	Create a new batch	CTRL+N
	Open the batch or add image to the batch	CTRL+O
	Export to file	CTRL+ALT+S
	Export to database	CTRL+ALT+D
	Undo an action	ALT+BKSP or CTRL+Z
	Redo an action	ALT+SHIFT+BKSP or CTRL+Y
	Cut the text and place it on the Clipboard	CTRL+X or SHIFT+DEL
	Copy the selection on the Clipboard	CTRL+INS or CTRL+C
	Insert Clipboard's contents in the text	CTRL+V or SHIFT+INS
	Delete selected text, selected pages, all blocks (after the command Select all)	DEL
	Delete pages (select components to delete)...	ALT+DEL
	Select the whole text in the current form field, or all pages of the batch, or all blocks on the current image	CTRL+A
	Find the necessary fragment in the recognized text	ALT+F3 or CTRL+F
	Repeat the search	F3
	Find the necessary fragment in the recognized text and replace it with another fragment	CTRL+H
	Specify a language for a fragment in the recognized field	CTRL+L
View menu	Display the whole image	CTRL+SHIFT+1
	Image width	CTRL+SHIFT+2
	Image height	CTRL+SHIFT+3
	Zoom in	CTRL+ Num *
	Zoom out	CTRL+ Num /
	Open the next batch page	CTRL+ Num +
	Open the previous batch page	CTRL+ Num –
	Open the page with a certain number	CTRL+G
	Update the page list	F5
	View the features of an active image (block)	ALT+ENTER
Scan&Read menu	Scan and recognize	F9
	Scan and recognize multiple pages	SHIFT+F9
	Open and recognize the image	CTRL+ALT+O
	Scan	CTRL+K
	Scan multiple pages	CTRL+SHIFT+K
	Stop scanning	CTRL +ALT+K
	Match template with selected pages	CTRL+E

	To:	Press:
Tools menu	Match template with all pages without blocks	CTRL+SHIFT+E
	Match template with a certain number	CTRL+ALT+E
	Recognize selected pages	CTRL+R
	Recognize all unrecognized pages	CTRL+SHIFT+R
	Recognize all unrecognized pages in the background mode	CTRL+ALT+R
	Open the "Templates" dialog	CTRL+ALT+T
	Check rules on selected pages	F6
	Check rules on all pages	SHIFT+F6
	Check recognition results	F7
	Check recognition results on all pages	SHIFT +F7
	Move to the next error	F4 (Num +)
	Move to the previous error	SHIFT+F4 (Num -)
	Move to the next error in the rule	F8
	Move to the previous error in the rule	SHIFT+F8
	Clear the error flag of the page	CTRL+Q
	Translate selected word or word combination	CTRL+SHIFT+T
	Open the "Language editor" dialog	CTRL+SHIFT+L
	Open the "Pattern editor" dialog	CTRL+SHIFT+A
	Open the "Scanner parameters" dialog	CTRL+SHIFT+S
	Open the "Options" dialog	CTRL+SHIFT+O
Window menu	Arrange windows	CTRL+W
Help menu	Get context help for the selected object (menu or dialog control)	SHIFT+F1
General commands	Open the selected page of the batch	ENTER
	Activate the Batch window	ALT+0
	Activate the Image window	ALT+8
	Activate the Form window	ALT+9
Hot keys of the Template editor		
	To:	Press:
Blocks menu	Save template	CTRL+S
	Undo	CTRL+Z
	Redo	CTRL+Y
	Delete selected blocks	DEL
	Delete blocks of a certain type	ALT+DEL
	Make the block "Text field"	CTRL+1
	Make the block "Checkmark"	CTRL+2
	Make the block "Radio group"	CTRL+3

	To:	Press:
Blocks menu	Make the block "Bar-code"	CTRL+4
	Make the block "Static text"	CTRL+5
	Make the block "Black line"	CTRL+6
	Make the block "Black square"	CTRL+7
View menu	Find block	CTRL+F
	Display the whole image	CTRL+SHIFT+1
	Image width	CTRL+SHIFT+2
	Image height	CTRL+SHIFT+3
	Zoom in	CTRL+ Num *
	Zoom out	CTRL+ Num /
	Display active block or image properties dialog	ALT+ENTER
Image menu	Select tool – Create a text field	ALT+1
	Select tool – Add a part of block	ALT+2
	Select tool – Delete a part of block	ALT+3
	Select tool – Move blocks	ALT+4
	Select tool – Renumber blocks	ALT+5
	Select tool – Delete blocks	ALT+6
	Open the "Language editor" dialog	CTRL+SHIFT+L
	Open the "Pattern editor" dialog	CTRL+SHIFT+A
	Open the "Options" dialog	CTRL+SHIFT+O
	Get context help for the selected object (menu or dialog control)	SHIFT+F1

ODBC Drivers

What is ODBC?

The ODBC protocol was developed by Microsoft so that applications could address databases, produced by different software companies, in a standard way.

ODBC has now become standard and is supported by nearly all databases.

ODBC Structure

ODBC consists of three parts: the application, the ODBC manager and the database driver. The ODBC manager was developed by Microsoft and is a part of the operating system. Database drivers are developed by independent software producers.

When an application addresses a database, the ODBC manager detects which driver operates this database and transmits the request to it. All real work with the database is done by its driver.

Installing ODBC

For normal ODBC work, you should install the ODBC manager and the database driver you want to address.

In Windows 95 and Windows NT the ODBC manager and the ODBC driver are supplied with the operating system and are installed on your computer automatically.

The database driver is installed after ODBC manager installation. Installation procedure depends on the database producer. Some products, e.g., Microsoft Office 7.0, install their drivers automatically. Others may require special operations to install them. If installed drivers do not include a driver for your database, refer to the documentation enclosed with this database.

What is a Data Source?

To address the database, you should give additional information to the database driver. The nature of the necessary information depends on the driver. For example, for the Paradox database you should indicate the name of the file, in which the database is located, and for the SQL Server you should indicate the name of the server, in which the database is located, the user's name, password, etc.

To save you specifying all this information each time you access the database, the ODBC manager retains the information in the "data source". When you next address the database it is sufficient to indicate the "data source", from which the driver can take all the necessary information.

Error Messages during Work with ODBC

If an error occurs during work with the database, a message is displayed which includes unfamiliar abbreviations. This is because all interaction with the data base is carried out by the ODBC driver and not by FineReader. When an error occurs it is the driver, and not FineReader, which displays an error message and gives additional technical information. In the documentation enclosed with the driver, there should be a complete check-list of such error messages with a detailed description of their reasons.

What is a Table in the Database?

When you address the ODBC from FineReader, selection of the data source is immediately followed by a request to select a table. “Table” means different things in different ODBC drivers.

For classic databases, such as Access or SQL Server, “table” means the table in the database.

When you work with the Excel driver “table” is a worksheet from the work book.

When you work with drivers for DBF and TXT files, “table” means a file. In this data source a catalog is specified, where the file is located. Each file represents one table.

When you select a table, you can either select one of the existing tables, or type the name of a new table. In the latter case, FineReader will automatically create a table with the specified name. Since different drivers impose different limitations on the names of tables, we advise you to use short names, consisting only of Latin letters and digits, beginning with letters and not containing spaces.

How to Achieve Export to a Database

- It is also necessary to install a driver for your database. The installation procedure is described in the documentation enclosed with your database.

You should create a new data source, giving all necessary information to the driver. This can be done either from the dialog box which selects the data source (the New button), or from the program of the ODBC manager configuration, found in the Control Panel.