# End-User License Agreement (EULA) For ABBYY Software

**IMPORTANT-READ CAREFULLY: This ABBYY End-User Licence Agreement ("EULA") is a legal agreement between you (either an individual or a single entity) and ABBYY Software House (ABBYY hereinafter), the copyright owner of the software (SOFTWARE hereinafter) supplied together with this agreement and may include associated media, printed materials, and "online" or electronic documentation. By installing, copying, or otherwise using the SOFTWARE, you agree to be bound by the terms of this EULA. If you do not agree to the terms of this EULA, do not install or use the SOFTWARE.**

## SOFTWARE LICENCE

This SOFTWARE is protected by copyright laws and international copyright treaties, as well as other intellectual property laws and treaties. This SOFTWARE is licensed, not sold.

### 1. GRANT OF LICENCE

This EULA grants you the following rights:

**1.1 Application Software**

You may install and use one copy of the SOFTWARE on a single computer.

**1.2 Portable Computer**

If you have purchased this SOFTWARE and are the primary user of this SOFTWARE you may copy this SOFTWARE onto your portable computer, providing the licence is not shared or used concurrently.

**1.3 Multiple Licence Package**

You may also store or install a copy of the SOFTWARE on a storage device, such as a network server, used only to install or run the SOFTWARE on your other computers over an internal network; however, you must acquire and dedicate a licence for each separate computer on which the SOFTWARE is installed or run from the storage device. A licence for the SOFTWARE may not be shared or used concurrently on different computers.

**1.4 Back-Up Copy**

You have the right to make a single copy of the software for back-up purposes in case the original SOFTWARE is lost or damaged. The back-up copy may not be used for any other purpose.

### 2. Copyright

All title and copyrights in and to the SOFTWARE (including but not limited to any images, photographs, animations, video, audio, music, text, and "applets" incorporated into the SOFTWARE), the accompanying printed materials, and any copies of the SOFT-WARE are owned by ABBYY or its suppliers. The SOFTWARE is protected by copyright laws and international treaty provisions. Therefore, you must treat the SOFTWARE like any other copyrighted material except that you may install the SOFTWARE on a single computer provided you keep the original solely for backup or archival purposes. You may not copy the printed materials accompanying the SOFTWARE.

### 3. Description of other Rights and Limitations

**3.1 Reverse Engineering, Decompilation, and Disassembly**

You may not disassemble, reverse engineer, or decompile (get the source code out of the object code) the SOFTWARE (applications, databases and other SOFTWARE components), except and only to the extent that such activity is expressly permitted by applicable law notwithstanding this limitation.

**3.2 Object Code**

You may not make changes to the object code of applications and databases other than those provided for by the SOFTWARE and described in the documentation

**3.3 Separation of Components**

The SOFTWARE is licensed as a single product. Its component parts may not be separated for use on more than one computer.

**3.4 Re-distribution of the SOFTWARE**

Any re-distribution of the SOFTWARE is strictly prohibited. Re-distribution includes granting of access to third parties (including network access and other access methods), and also selling, renting, leasing or lending of the SOFTWARE.

**3.5 Transfer of Rights**

You may not transfer the SOFTWARE copyright and other rights related to the SOFTWARE to any other person or entity.

### 4. Support Services

ABBYY may provide you with support services related to the SOFTWARE ("Support Services"). Use of Support Services is governed by the ABBYY policies and programs described in the user manual, in "online" documentation, and/or in other ABBYY-provided materials. Any supplemental software code provided to you as part of the Support Services shall be considered part of the SOFT-WARE and subject to the terms and conditions of this EULA. With respect to technical information you provide to ABBYY as part of the Support Services, ABBYY may use such information for its business purposes, including product support and development. ABBYY will not utilize such technical information in a form that personally identifies you.

## 5. Not for Resale Software.

If the SOFTWARE is labelled "Not for Resale" or "NFR," then, notwithstanding other sections of this EULA, you may not resell, or otherwise transfer against value, the SOFTWARE.

## 6. Upgrades

If the SOFTWARE is labelled as an upgrade, you must be properly licensed to use a product identified by ABBYY as being eligible for the upgrade in order to use the SOFTWARE. The SOFTWARE labelled as an upgrade replaces and/or supplements the product that formed the basis for your eligibility for the upgrade. You may use the resulting upgraded product only in accordance with the terms of this EULA. If the SOFTWARE is an upgrade of a component of a package of software programs that you licensed as a single product, the SOFTWARE may be used and transferred only as part of that single product package and may not be separated for use on more than one computer.

## 7. Dual-Media SOFTWARE

You may 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 for your single computer. You may not use or install the other medium on another computer. You may not loan, rent, lease, or otherwise transfer the SOFTWARE on the other medium to another user.

## 8. Termination

Without prejudice to any other rights, ABBYY may terminate this EULA if you fail to comply with the terms and conditions of this EULA. In such event, you must destroy all copies of the SOFTWARE and all of its component parts.

## 9. Valid Legislation

This EULA for SOFTWARE is subject to the country law in which it was purchased.

## 10. Limited Warranty

ABBYY warrants that (a) the SOFTWARE will perform substantially in accordance with the accompanying written materials for a period of ninety (90) days from the date of receipt, and (b) any Support Services provided by ABBYY shall be substantially as described in applicable written materials provided to you by ABBYY, and ABBYY support engineers will make commercially reasonable efforts to solve any problem issues. Some states and jurisdictions do not allow limitations on duration of an implied warranty, so the above limitation may not apply to you. To the extent allowed by applicable law, implied warranties on the SOFTWARE, if any, are limited to ninety (90) days.

## 11. Customer Remedies

ABBYY's and its suppliers' entire liability and your exclusive remedy shall be, at ABBYY's option, either (a) return of the price paid, if any, or (b) repair or replacement of the SOFTWARE that does not meet ABBYY's Limited Warranty and which is returned to ABBYY with a copy of your receipt. This Limited Warranty is void if failure of the SOFTWARE has resulted from accident, abuse, or misapplication. Any replacement SOFTWARE will be warranted for the remainder of the original warranty period or thirty (30) days, whichever is longer. Outside the United States, neither these remedies nor any product support services offered by ABBYY are available without proof of purchase from an official stockist.

## 12. No other Warranties

To the maximum extent permitted by applicable law, ABBYY and its suppliers disclaim all other warranties and conditions, either express or implied, including, but not limited to, implied warranties of merchantability, fitness for a particular purpose, title, and non-infringement, with regard to the SOFTWARE, and the provision of or failure to provide Support Services. This limited warranty gives you specific legal rights. You may have others, which vary from state/jurisdiction to state/jurisdiction.

## 13. Limitation of Liability

To the maximum extent permitted by applicable law, in no event shall ABBYY or its suppliers be liable for any special, incidental, indirect, or consequential damages whatsoever (including, without limitation, damages for loss of business profits, business interruption, loss of business information, or any other pecuniary loss) arising out of the use of or inability to use the SOFTWARE or the provision of or failure to provide Support Services, even if ABBYY has been advised of the possibility of such damages. In any case, ABBYY's entire liability under any provision of this EULA shall be limited to the greater of the amount actually paid by you for the SOFTWARE or US Dollar 5.00; provided, however, if you have entered into a ABBYY Support Services Agreement, ABBYY's entire liability regarding Support Services shall be governed by the terms of that agreement. Because some states and jurisdictions do not allow the exclusion or limitation of liability, the above limitation may not apply to you.

## 14. No Warranty

If you acquired THE SOFTWARE product VIA WEB DOWNLOAD, the following warranty EXCLUSION applies to you:
ABBYY expressly disclaims any warranty for the SOFTWARE. The SOFTWARE and any related documentation is provided "as is" without Warranty or condition of any kind, either express or implied, including, without limitation, the implied warranties and conditions of merchantability, fitness for a particular purpose, or non-infringement. The entire risk arising out of use or performance of the SOFTWARE remains entirely with you.

**Optical Character Recognition Application**

# ABBYY FineReader

**Version 5.0**
**User's guide**

# Contents

# Welcome!

Thank you for choosing ABBYY FineReader!

We all need to input texts into computer. It could be text of newspaper and magazine articles, contracts, business letters, faxes, price lists, question-naires, etc. 90% of all the documents we deal with are printed on paper. For years on there was only one way to input printed documents - you've had to type them in from the keyboard. Remember the long hours you've spent typing text from a document? What a great thing it could be if the computer could read the text by itself, straight from the sheet of paper!

Sometimes dreams come true. The FineReader Optical Character Recognition (OCR) application will allow your computer and scanner to do just that - to read printed text by themselves.

*But can't the scanner do the job on its own?*

No. The scanner only makes a picture, a photograph of the text. A photo-graph is a set of black and white dots, or, in other words, *an image file*, and it can't be edited in text editors such as MS Word, WordPerfect, Word Pro, etc. You need an OCR system that will look for symbols in the set of black and white dots, "recognize" letters in each symbol and turn the image into computer-editable text. Such an application will create a file that text edi-tors and desktop systems can cope with.

*So now I can input my document to my computer automatically?*

Yes, now you can input your document to your computer automatically without retyping it on the keyboard.

Enjoy!

## The User's manual

This user' guide introduces you to the basics of using of ABBYY FineReader. Every chapter starts with its short description and the chapter contents.

## Online Help

FineReader's online Help contains basic and advanced information on features, settings and dialogs. The online Help is provided in HTML format and has been designed for quick and easy information retrieval.

## Readme file

The Readme file contains last-minute information about the software.

## Technical Support

If you have any questions on how to use FineReader, please consult all the documentation you have (the User's Guide and the Help File) before contacting our technical support service. You may also browse the technical support section on our website www.abbyy.com. You may find the answer to your question there.

In case you didn't manage to find the answer, e-mail us at support@abbyy.com. Our technical support service experts will need the following information to answer your question competently:

- The serial number of your FineReader distributive.

- Model of your scanner.

- General description of your problem and the full error message text (if there was one).

- The version of your Windows operating system.

- You may supply us with any other information you consider important.

You can get some information automatically by clicking the **System Info** button in the **About ABBYY FineReader** dialog (menu **Help/About**).

C h a p t e r   1

# Installing and Starting ABBYY FineReader

This chapter deals with ABBYY FineReader installation procedures and related subjects, such as system requirements and workstation and network installations.

A special installation program supplied in the distribution package carries out the set up of FineReader. Always use the diskette or the CD-ROM supplied in the distribution package you got when you bought FineReader. Setup and CD-ROM files are saved in such a way that the installation program will work incorrectly if you copy these files to another diskette or CD-ROM.

## Chapter contents:

- Software and hardware requirements
- Installing ABBYY FineReader
- Installation on a network server and on a network workstation
- Starting ABBYY FineReader

# Software and Hardware Requirements

## You will need the following to run ABBYY FineReader:

1. PC with a Intel® Pentium® 133MHz processor or higher.

2. Microsoft® Windows® 2000, Windows® NT® Workstation 4.0 with Service Pack 3 or greater, Windows® 95, or Windows® 98.

3. 32 Mb of RAM, plus 16 Mb of RAM for each additional processor (on a multiprocessor system)

4. Microsoft® Internet Explorer 3.02 or higher (Microsoft® Internet Explorer 5.01 included in the distribution package).

5. 40 Mbytes of free hard-disk space for minimal program installation.

6. 50 Mbytes of free hard-disk space for the program operation.

7. 100% Twain-compatible scanner, digital camera or fax-modem.

8. CD-ROM drive.

9. 3.5" disk drive (for CD+diskette setup media).

10. Mouse or other pointing device.

11. VGA or other high-resolution monitor.

# Installing ABBYY FineReader

## Installation options

After the setup program has checked the system, type in your name and indicate the folder you want to install FineReader in. The setup program will then display several installation options. Select the option you like by clicking the appropriate button.

**Full** – all the components are installed including all recognition languages.

**Minimal** (recommended) – the application is installed in its minimum configuration: a single interface language selected during installation, and two recognition languages – English plus one more selected during installation.

**Custom installation** – you may choose any number of components included in the distribution package (including available recognition languages).

---

**N o t e :** If you want to use your user dictionaries and user patterns from the previous version, don't uninstall it prior to the new version installation. After you've installed the new version, the **patterns and dictionaries of your previous version** will be available in the new version.

We recommend that you close all applications (e.g. Microsoft Outlook, Microsoft Word etc.) before running the ABBYY FineReader setup program.

### Installing ABBYY FineReader

**If there are both a CD-ROM and a diskette in the distribution package, do the following:**

1. Insert the "start-up" diskette in the floppy disk drive.

2. Insert the CD-ROM into the CD-ROM drive.

3. Click the **Start** button on the **Taskbar** and select the **Settings/Control Panel** item.

4. Double-click the **Add/Remove Programs** icon.

5. Select the **Install/Uninstall** tab and click the **Install** button.

6. Follow the installation instructions.

**N o t e :** Steps 3-5 have the same effect as running the **install.exe** program from the "start-up" distribution diskette.

**If there is only a CD-ROM in the distribution package, do the following:**

1. Insert the CD-ROM into the CD-ROM drive.

2. Click the **Start** button on the **Taskbar** and select the **Settings/Control Panel** item.

3. Double-click the **Add/Remove Programs** icon.

4. Select the **Install/Uninstall** tab and click the **Install** button.

5. Follow the installation instructions.

**N o t e :** Steps 2-4 have the same effect as running the install.exe program from the CD-ROM.

# Installation on a Network Server and on a Network Workstation

**Installation on a Network Server (for system administrators only)**

Installation of FineReader 5.0 Office on a network server can only be performed by system administrator. To do it:

- If there are both a CD-ROM and a diskette in the distribution package, run Install.exe from the \Server folder on the start-up diskette.

- If there is only a CD-ROM in the distribution package, run Install.exe from the \Server folder on the CD-ROM.

## Additional licenses

If FineReader is used by more than one user, at the end of installation on a network server you need to add serial numbers for a correct network processing. To do it:

1. Run LicSetup.exe from the folder where FineReader 5.0 Office was installed. The **Add License** dialog will be displayed.

2. Input a new serial number and click the **Add** button.

### Note:
1. You cannot use logical drives created by the SUBST command.
2. Such components as MS Windows NT4.0 Service Pack 3.0 or MS Internet Explorer 5.01 won't be installed if you choose installation on a network server. To avoid difficulties related to the absence of these components, consult a system administrator so that he updates the network workstation according to system requirements before the installation.
3. Check before installation, that all users have read-write permissions for the network folder where FineReader 5.0 Office is being installed.

## Installation on a Network Workstation

If FineReader 5.0 Office was installed on a network server, you can run the setup program from that server – you don't have to use the CD-ROM or/and diskette. But if you would like to install FineReader on the server in the workstation installation mode, the installation diskette is necessary.

**To install ABBYY FineReader 5.0 Office on a workstation:**

- Run NetSetup.exe from the network folder where
  FineReader 5.0 Office has been installed.
  Follow the installation instructions.

### Note:
1. You should have administrative rights for the workstation where FineReader is being installed.
2. If the message "Can't load FineReader. There is no free license." is displayed, check the number of additional licenses you have added in the **Add license** dialog and the number of users that are working with FineReader at this moment.

# Starting ABBYY FineReader

### To start ABBYY FineReader:

- Select the **ABBYY FineReader 5.0 Pro (Office)** item in the **Start/Programs/ABBYY FineReader 5.0** menu.

---

**N o t e :**  Make sure your scanner is connected to the computer and plugged-in and powered-up before you start FineReader. If your scanner has not been connected yet, please consult the user guide supplied with the scanner for instructions on scanner installation.

---

If you don't have a scanner, FineReader may recognize your image files (see a sample **demo.tif** file in the **ABBYY FineReader/Demo** folder).

Chapter 2
# Quick Start

In this chapter you will learn how to input a document without even knowing anything about how ABBYY FineReader works!

You will also learn what windows and toolbars are there in FineReader. If you're already experienced with FineReader you may skip all text up to "New features of ABBYY FineReader 5.0" section.

## Chapter contents:

- How to input a document in a minute
- The ABBYY FineReader main window
- ABBYY FineReader toolbars

# How to Input a Document in a Minute

1. Turn on the scanner in case it has a separate power source from your PC.

---

**Note:** Many scanner models have to be turned on before you turn on the computer.

---

2. Turn on the computer and start FineReader (**Start/Programs/ABBYY FineReader 5.0/ABBYY FineReader 5.0 Pro** or **Office**). You will see the FineReader main window.

3. Put the page you want to read into the scanner.

4. Click the arrow right to the **Scan&Read** button.
   Select the **Scan&Read Wizard** item in the local menu.



**Scan&Read Wizard** is a special scan&read/open&read mode. You can find a sample image file in the **Default** folder. This folder is located in the folder you have installed FineReader in. FineReader will prompt you at every necessary step.

5. Follow the **Scan&Read Wizard** instructions.

Document input process is made up of four steps: *scanning, reading, spelling check* and *saving* the recognized text.

After the scanning is done, a "photograph" of the source page will appear in the **Image** window. Then the application asks you to set the recognition parameters. After you are done setting them, it starts recognizing the image analyzing its layout at the same time. Already processed image areas are painted blue.

The recognized text will be displayed in the **Text** window. You can check and edit this text in this window. After you are done checking the document, **Scan&Read Wizard** will prompt you to either send the recognized text to the application you specify, or save it to a file, or go on processing more images.

# The ABBYY FineReader Main Window

FineReader performs all document processing in batch mode. A batch is a folder containing images, recognized text files and other FineReader information files. Every scanned image is made a separate batch page. If there are several images in a single image file (for example, if you are dealing with a multipage TIFF), each image of this file will be made a separate batch page.

When you start FineReader for the first time, it opens the default batch. There is only one page in this batch, a file with the **demo.tif** image. You may work with the default batch or create a new batch of your own. See "General Information on Working with Batches" section for more details.



**Main Window**

**Standard Toolbar**

**Formatting Toolbar**

**WizardBar**
provides tools for full text processing: Scanning, Recognition, Spelling Check and Saving

**Text Window**
displays the recognized text for checking and editing

**Image Window**
displays the scanned image for viewing and drawing blocks

**Zoom Window**
displays the zoomed-in image of the text line you edit or part of an image you are working on

**Batch Window**
displays the pages of the open batch in one of two modes: thumbnails (as now) or details

**Image Tools Toolbar**
provides tools for drawing and editing blocks, zoom tools and tools for image editing

You will see the FineReader main menu at the top of the FineReader main window. The following four toolbars are displayed under the main menu: the **Standard**, the **Formatting**, the **Image Tools**, and the **WizardBar** toolbars. You may show/hide any toolbar.

You can show/hide a toolbar by clicking the **Toolbar** item in the **View** menu or via the local menu. Right-click any toolbar to open the local menu. You will see the toolbar list, the currently shown toolbars are selected. Click the toolbar name you want to show/hide.

At the bottom of the FineReader main window there is a status bar which displays information about the application status and the operations it is currently performing, as well as brief information on the menu items and buttons you've selected.

The **Batch** window is always displayed in the **Main window**. Three more windows may also be displayed: the **Image**, the **Zoom** and the **Text** windows.

By default, the **Zoom** window displays a black and white image even if you open a color or a gray image. If you want the **Zoom** window to display your

color images in color, clear the **Show black and white images in Zoom window** item in the **Image window** group on the **View** tab (menu **Tools>Options**).

You may change the windows arrangement on the screen:

- Select one of the following items: the **Batch Window** >...; **Image and Text Windows** >...; **Zoom Window** >.... in the **View** menu.

| Some recommended windows arrangements: | Useful if/when: |
|---|---|
| The **Batch** window on the left; **Batch View**: **Thumbnails**; the **Image**, **Text** and **Zoom** windows | there aren't too many pages in the batch |
| The **Batch** window at the top; **Batch View**: **Details**; the **Image**, **Text** and **Zoom** windows | there are many pages in the batch |
| The **Batch** window at the top; **Batch View**: **Details**; the **Image** and **Zoom** windows | you perform layout analysis and recognition |
| The **Batch** window at the top; **Batch View**: **Details**; the **Text** and **Zoom** windows | you edit the recognized text |

**To switch between windows:**

- Press **CTRL+TAB.**
- Press **ALT+1** to activate the **Batch** window.
- Press **ALT+2** to activate the **Image** window.
- Press **ALT+3** to activate the **Text** window.

# ABBYY FineReader Toolbars

There are four toolbars in FineReader: the **Standard**, **Image Tools**, **Formatting** and **WizardBar** toolbars. The toolbar buttons are definitely the most convenient way to access the application functions; however, the same functions may be accessed via menus or hot keys. If you want to find out what a particular toolbar button is for, just move the mouse pointer to it. The button's tooltip will be displayed, and the status bar will display some more button details.

## The WizardBar toolbar
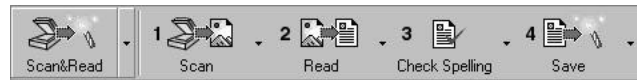
The **WizardBar** toolbar buttons launch the main FineReader functions: **Scanning, Reading, Checking** and **Saving** the recognition results. The numbers on the buttons indicate the order in which the respective document input actions should be performed. You may perform each action separately or you may combine them into one by clicking the **Scan&Read Wizard** button. The **Scan&Read Wizard** performs the full document processing cycle automatically.

Each button features several function modes. Click the arrow right to the button and select the mode you like in the local menu. The button icon always displays the last selected mode. Click the button itself to run this mode again.

**Scan&Read**

**Scan&Read Wizard** – runs a special scan&read mode. FineReader guides you through the document processing process and advises you what you should do to get this or that result.

**Scan&Read** – starts scanning and reading a document using the current options.

**Scan&Read Multiple Images** – scans and reads several images in a row.

**Open&Read** – opens and reads the images selected in the **Open** dialog.

**1-Scan**

**Open Image** – adds image(s) to the batch. Each added image is copied to the batch folder.

**Scan Image** – scans an image.

**Scan Multiple Images** – scans images until you stop the scanning. Select the **Stop Scanning** item in the **File** menu to stop the scanning.

**Options** – opens the **Scan/Open Image** tab (**Options** dialog), so you may set scanning options etc.

**2-Read**

**Read** – reads the open batch page.

**Read All** – reads all unrecognized batch pages.

**Options** – opens the **Recognition** tab (**Options** dialog) so you can set the document recognition options.

---

**3-Check Spelling**
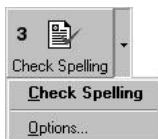
**Check Spelling** – searches the text for misspelt and uncertain words (the ones with uncertainly recognized characters).

**Options** – opens the **Check spelling** tab (**Options** dialog) so you can set the spelling check options.

---

**4-Save**
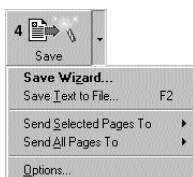
**Save Wizard** – opens the **Save Wizard** so you may select the application for saving the text and also set the saving options.

**Save Text to File** – saves the recognized text to a disk file.

**Send Selected Pages to** – you may only want to export the selected batch pages. In this case select the pages you want to export, then select this item and then select the application you want to export your pages to. FineReader will export your pages to the application you select without saving the text on disk.

**Send All Pages to** – exports all recognized pages into the application you select without saving the text on disk.

**Options** – opens the **Formatting** tab (**Options** dialog) so you may set the saving options.

## The Formatting toolbar

The **Formatting** toolbar buttons feature various text formatting tools. You may edit the text and text formatting in the **Text** window.

Font    Font size    Italic    Superscript    Display nonprinted characters

Bold    Underline    Subscript    Show paragraph formatting
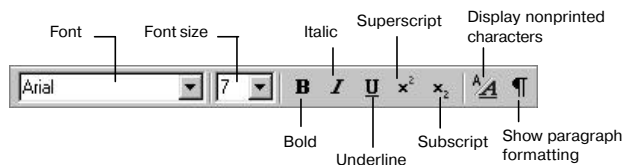
## The Image Tools bar

The **Image Tools** bar buttons feature page layout analysis (e.g. creating and editing blocks) tools, as well as tools for increasing/decreasing the image scale and image editing (e.g. image despeckle etc.).

Analyze layout

Draw block tools
- Draw recognition area
- Draw text block
- Draw table block
- Draw picture block
- Select objects

Block frame and position tools
- Add block part
- Cut block part
- Renumber blocks
- Delete blocks

Table block tools
- Add vertical separator
- Add horizontal separator
- Delete separator

Image tools
- Zoom Out
- Zoom In
- Eraser
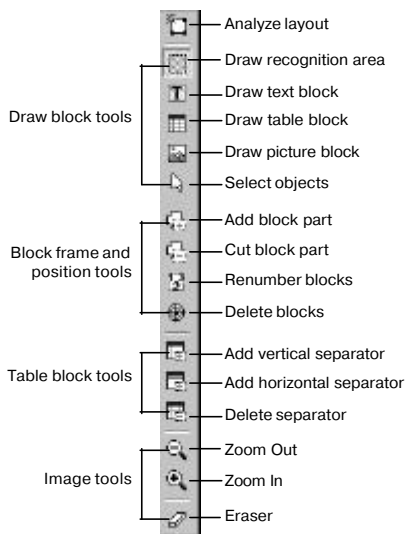
## The Standard toolbar

The **Standard** toolbar buttons feature file and image tools (undo/redo an action, scrolling the batch pages, clean and rotate image) and the Recognition languages list as well.



---
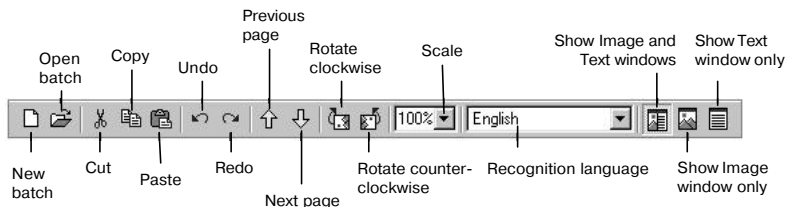
**Note:** The view of FineReader main window, or, to be more precise, the number of buttons displayed on FineReader toolbars, depends on your monitor's resolution. To display all the available buttons increase your monitor's resolution. However, the FineReader functionality is not reduced if some buttons are not displayed - the buttons feature only a part of available FineReader functions, and there is always a menu item for each button.

# General Features of ABBYY FineReader

ABBYY FineReader provides you with the for document input into the computer. All you have to do is click the **Scan&Read** button once – all the rest is done by the application, so you don't have to spend any time studying its intricacies. You can send the recognized text to a word processor or a spreadsheet application, you can save it in RTF/DOC, PDF and HTML thus retaining the full document layout, or you can export the recognized text into a database.

## Chapter contents:

- What is an OCR-system?
- New features of ABBYY FineReader 5.0
- Supported document saving formats
- Supported image formats

# What is an OCR System?

An OCR (Optical Character Recognition) system enables you to input printed documents into your computer automatically via a scanner.

FineReader is an omnifont optical text recognition system. It means that the system recognizes texts set in practically any font without prior training. One of FineReader's characteristic features is its high recognition accuracy and low sensitivity to print defects. These features are due to a special recognition technology based on the principles of Integral Purposeful Adaptive (IPA) perception which is fully implemented in FineReader.

**The process of document input can be divided into two stages:**

1. **Scanning.** During the first stage the scanner acts as the "eye" of your computer: it looks at the image and transfers it into the computer. The acquired image is nothing more than a picture, a set of black, white and color dots impossible to edit in any word processor.

2. **Recognition.** During the second stage FineReader performs the OCR image processing.

**Let's take a closer look at second stage.**

FineReader OCR image processing involves analyzing the image file transmitted by the scanner (layout analysis) and recognizing each character. The processes of layout analysis (selecting the recognition areas, tables, pictures, lines and separate characters) and reading the image interact closely: the layout analysis procedure uses the information about the recognized text for a more accurate page layout analysis.

As mentioned above, the image recognition process is based on the principles of Integral Purposeful Adaptive perception.

- **Integrity** – the object of recognition is described as a single entity by a set of basic elements and their interrelations.

- **Purposefulness** – recognition is a process of generation and purposeful verification of hypotheses.

- **Adaptivity** – the system's ability to learn and be trained.

These three principles determine the system's behavior. The system generates a hypothesis about an object of recognition (a character, a part of a character, several glued characters) and then accepts or rejects it, trying to find all the structural elements and their interrelations step by step. These structural elements are computer equivalents of character parts crucial for human perception (arcs, circles, dots etc.). Then the application adapts itself to the text – tunes itself – using the positive experience gained from the first confidently recognized characters. Purposeful searching and using context

information enable the system to recognize even torn and distorted charac‐
ters, so it is nearly insensitive to print defects.

The final result is the recognized text you see in the FineReader **Text** win‐
dow, a text you can edit and save in any convenient format.

# New Features of ABBYY FineReader 5.0

### Working with images

- New *thumbnail* mode of batch page display. In this mode the batch
  pages are displayed as thumbnails — very small page images.

### Scanning

- Dual page processing. Two pages of an open book are scanned simulta‐
  neously but processed separately. A single dual page image is split into
  two different batch pages.

### Page Layout Analysis

- ABBYY FineReader 5.0 features a new procedure of page layout analy‐
  sis. Now the page layout analysis is coupled with recognition, so that
  the page layout analysis procedure uses the information about the rec‐
  ognized text for more accurate page layout analysis. This new procedure
  of layout analysis is used automatically when the application recognizes
  a page, so the layout analysis and recognition are both done simultane‐
  ously. The old layout analysis procedure, the one that doesn't use recog‐
  nition information, is still available (see menu **Process>Analyze Layout**).

- "Embedded" pictures recognition — you can draw a "picture" block
  inside a text or a table block.

### Recognition

- EU version with 121 recognition languages (ABBYY FineReader
  EU version).

- Cyrillic Plus version with 55 additional Cyrillic languages,
  total 176 languages (ABBYY FineReader Cyrillic Plus version).

- Support of Armenian alphabet (in both EU and Cyrillic Plus versions).
  See the full list of supported languages in the "Supported Languages"
  section of ABBYY FineReader Help.

- Formal languages (Basic, Cobol, Fortran, Java, C++, Pascal)
  recognition.

- Subscript characters recognition (e.g. $H_2O$).

- Vertical text recognition.

- Full text color retention.

### Saving and Editing

- Blocks and the recognized text are synchronized: if you delete a block, the corresponding recognized text section is deleted; if your renumber the blocks, the text in the **Text** window is rearranged accordingly.

- Saving in the HTML format with full page layout retention.

- Saving in RTF, PDF and HTML formats with full text color retention.

- Saving in RTF, DOC, XLS, HTML, TXT and CSV formats with full Unicode encoding support.

# Supported Document Saving Formats

### ABBYY FineReader saves the recognition results in the following formats:

- Microsoft Word Document(*.**DOC**)

- Rich Text Format (*.**RTF**)

- Adobe Acrobat Format (*.**PDF**)

- **HTML**

- Comma Separated Values file (*.**CSV**)

- Plain Text (*.**TXT**). ABBYY FineReader supports various code pages (Windows, DOS, Mac, ISO) and the Unicode encoding.

- Microsoft Excel Spreadsheet (*.**XLS**)

- **DBF**

# Supported Image Formats

### ABBYY FineReader opens image files in the following formats:
**BMP:**

2-bit — black and white
4- and 8-bit — Palette

16-bit
24-bit − Palette and TrueColor
32-bit

**PCX, DCX:**

2-bit − black and white
4- and 8-bit - gray

**JPEG:**

gray and TrueColor

**TIFF:**

black and white − uncompressed, CCITT3, CCITT3FAX, CCITT4, Packbits
gray − uncompressed, Packbits, JPEG
TrueColor − uncompressed, JPEG
Palette − uncompressed, Packbits
multi image TIFF

**PNG:**

black and white, gray, color

## ABBYY FineReader saves image files in the following formats:

**BMP:**

black and white, gray, color

**PCX:**

black and white, gray

**JPEG:**

gray, color

**TIFF:**

black and white − uncompressed, CCITT3, CCITT3FAX, CCITT4, Packbits
gray − uncompressed, Packbits, JPEG
color − uncompressed and JPEG

**PNG:**

black and white, gray, color

C h a p t e r  4

# Acquiring the Image

Recognition quality depends greatly on the quality of the source image. In this chapter you will learn how to scan the document properly, how to open and read an image you already have on your PC (see the list of supported image formats in "Supported Image Formats" section), how to process images and make them good for recognition (eliminate scanning "dust" and so on) etc.

## Chapter contents:

- Scanning
- Setting scanning parameters
- Tips on brightness tuning
- Scanning multi-page documents
- Scanning dual pages
- Opening images
- Working with image
- Page numbering
- Batch image options

# Scanning

FineReader "talks" with scanners via the TWAIN interface.
It is a universal standard adopted in 1992 in order to unify the interaction
of devices providing image input into computer (such as scanners) and
external applications. There are two ways FineReader can "talk" with
a scanner via a TWAIN-driver:

- **using its own interface:** in this case you will use the FineReader
  **Scanner settings** dialog to set scanning options; the respective option
  is **Use FineReader Interface**;

- **using scanner's TWAIN interface:** in this case you will use the scanner's
  TWAIN-dialog to set the scanning options; the respective option is
  **Use TWAIN-Source Interface.**

**Both modes have their advantages and drawbacks**

When you select the **Use TWAIN-Source** interface option, the preview
image option usually becomes available. The preview lets you set the scan-
ning area and tune the brightness precisely - it is done easily because you
can see how your changes affect the previewed image. A certain drawback is
that different scanners have different TWAIN driver dialogs. The instructions
on how to use the scanner's TWAIN-dialog are supplied with the scanner.

When you select the **Use FineReader interface** option, the following fea-
tures become available: first, you may scan multiple images on scanners
without ADF, second, you may save scanning options in batch template file
(\*.fbt) and then use them for other batches.

You may easily switch from one mode to the other:

- Select the **Scan/Open Image** tab in the **Options** dialog (menu
  **Tools>Options**) and click the radio button you want - either
  **Use TWAIN-Source interface** or **Use FineReader interface.**

---

**N o t e :**
1. You may see the **Use FineReader interface** option unavailable (dis-
   abled) if you use certain scanner models.
2. If you want to see the **Scanner Settings** dialog in the **Use FineReader
   interface** mode, select the **Display options dialog before scanning** item
   on the **Scan/Open Image** tab (**Tools>Options**).

---

**Important:** Consult the documentation supplied with the scanner to
connect it correctly. After you've connected the scanner to the computer,
do not forget to install a TWAIN-driver and/or a scanning application.

## To start scanning:

Click the **1-Scan** button or select the **Scan** item in the **File** menu.
The **Image** window displaying a "photograph" of the scanned page will
appear in the FineReader main window. If you want to scan several pages,
click the arrow right to the **1-Scan** button and select the **Scan Multiple
Images** item.

If scanning doesn't start right away, a dialog will open. It'll be one of the fol-
lowing two:

- The scanner's **TWAIN-Source** dialog. Check the scanning options
  and click the **OK** button to start scanning.

- The **Scanner Settings** dialog. Check the scanning options and click
  the **OK** button to start scanning.

**Tip:** If you want to start recognition immediately after the source images
are scanned, you may use the **Scan&Read** or **Scan&Read Multiple Images**
option:
- Click the arrow right to the **Scan&Read** button and select either
**Scan&Read** or **Scan&Read Multiple Images** item in the local menu.
FineReader will scan and read the images. The **Image** window displaying
a "photograph" of the scanned page and the **Text** window displaying the
recognition results will appear in the FineReader main window. The recog-
nized text may be sent to various external applications and saved in various
formats.

# Setting Scanning Parameters

Recognition quality depends greatly on the scanned image quality. The
image quality may be adjusted by setting the main scanning parameters: res-
olution, scan mode and brightness.

The main scanning parameters are:

- **Resolution** — use 300 dpi resolution for regular texts (font size 10pt.
  or greater) and 400-600 dpi resolution for texts set in smaller font sizes
  (9pt. or less).

- **Scan mode — gray.**
  Scanning in grayscale mode is best for recognition. If you scan your
  images in grayscale, the application tunes the brightness automatically.

- **Scan mode — black and white.**
  Black and white scan mode enables the system to scan at a higher scan-
  ning speed, but at the same time some character information is lost.
  This may have a negative effect on the recognition quality of the docu-
  ments of medium and low print quality.

- **Scan mode —  color.**
  If your scan color documents where there are pictures, colored charac-
  ters and background — you may want to keep the color in your electro-
  nic document. Use the **color** scan mode in this case. Otherwise use
  **gray** scan mode.

- **Brightness** – in most cases the medium brightness value (50%) will do.
  Some documents scanned in black and white mode may require some
  additional brightness tuning.

---

**N o t e :** Scanning at 400-600 dpi resolution (instead of the default 300 dpi)
or scanning in gray or color (instead of black& white) mode takes more
time. In case of certain scanner models 600 dpi resolution scanning takes up
to four times more time than 300 dpi resolution scanning.

---

**To set scanning parameters:**

- If you want to scan your images using the **FineReader TWAIN inter-
  face**, select the **Scanner Settings** item in the **Tools** menu. The **Scan-
  ner Settings** dialog will open. Set the scanning options in the dialog.

- If you want to scan your images using the **TWAIN-Source interface**,
  your scanner's TWAIN dialog opens automatically when you click the
  **1-Scan** button. Set the scanning parameters in the dialog. Scanning
  options may have different names depending on the scanner model;
  for example, the word "threshold" may be used for brightness, or the
  brightness control may be displayed as a "sun" or a black and white
  circle. The exact options are described in the documents supplied
  with your scanner.

# Tips on Brightness Tuning

The scanned image has to be legible. To check if it is legible, view the image in the **Zoom** window.

**brightness** – an example of a good image (from the OCR point of view)

If you see that the scanned image is far from perfect (characters are glued or torn), consult the table below to find how you can improve the image quality.

| Your image looks like this: | Possible remedy: |
|---|---|
| brightness  characters are "torn" or very light | <ul><li>Try to decrease the brightness (it will make the image darker).</li><li>Try to scan it in gray mode (the brightness autotuning is used in this case).</li></ul> |
| brightness  characters are distorted, glued, or filled | <ul><li>Try to increase the brightness (it will make the image brighter).</li><li>Try to scan it in gray mode (the brightness autotuning is used in this case).</li></ul> |

# Scanning Multi-page Documents

FineReader supports a special scanning mode convenient for scanning multi-page documents: **Scan Multiple Images**. You may scan as many pages as you want in this mode. However this mode does have some peculiarities:

- If you scan your images using the **FineReader TWAIN interface,** the scanner doesn't stop after it is finished scanning one page, it automatically starts scanning the next one.

- If you scan your images using the **TWAIN-Source interface,** the **TWAIN-dialog** of the scanner doesn't close after the scanner is finished scanning a page. You can put the next page into the scanner and scan it, then another, and so on.

If you have to scan many pages, there are two ways to do it: you may use a scanner with an Automatic Document Feeder (ADF) or one without!

**Scanning with ADF:**

1. In case you are using the **FineReader interface**, select the **Use ADF** option in the **Scanner Settings** dialog (menu **Tools>Scanner Settings**) and then select **File>Scan Multiple Images** menu item to start scanning multiple images.

2. In case you are using the **TWAIN-Source interface**, select the **Use ADF** option in the TWAIN-dialog of your scanner (keep in mind that this option may be named differently depending on the scanner model; please consult the documents supplied with your scanner for the exact procedure) and then select **File>Scan Multiple Images** menu item to start scanning.

**Scanning without ADF:**

1. In case you are using the **FineReader interface**

- Select the **Scan Multiple Images** item in the **File** menu

There are certain ways to increase the convenience of scanning several pages using a flatbed scanner without ADF. You may make use of the following two alternatives :

- Set the pause value – the time elapsing between the moment the scanner finishes scanning one page and the moment it starts scanning the next one. Select the **Pause between pages** option and then set the pause value (in seconds) in the **Scanner Settings** dialog (menu **Tools>Scanner Settings…**).

   In this mode, the scanner doesn't begin scanning the next page until the specified number of seconds has elapsed, so you have sufficient time to put the next page into the scanner. After this pause, the scanning continues automatically.

- Select the **Stop between pages** option in the **Scanner Settings** dialog (menu **Tools>Scanner Settings**).

   In this case every time the scanner completes scanning a page a dialog asking you if you want to continue scanning will be displayed. Click the **Yes** button to continue scanning or **No** to finish scanning.

When you are finished scanning your pages, select the **Stop Scanning** item in the **File** menu.

2. In case you are using the **TWAIN-Source interface**

- Select the **Scan Multiple Images** item in the **File** menu. The TWAIN dialog of your scanner will open. Click the **Scan** (or **Final**, or what the actual name of this button is – it depends on your scanner's TWAIN dialog) button to start scanning.

Scan your page, put another page in your scanner and click the **Scan** button in the **TWAIN-dialog** of your scanner to scan another and so on.

When you are finished scanning your pages, click the **Close** (or what this button is actually named in the particular TWAIN dialog of your scanner) button in the **TWAIN-dialog** of your scanner..

---

**Tip:** You may want to control the quality of your scanned images more directly. To do this, select the **Open image during scanning** option on the **Scan/Open Image** tab (menu **Tools>Options**). If you select this option, each scanned page will be opened in the **Image** window immediately after it has been scanned. If you think the image is scanned incorrectly, stop the scanning (select the **Stop Scanning** item in the **File** menu) and re-scan the image.

---

# Scanning Dual Pages

When you scan a book, you'll find it more convenient to scan both the left and right book pages (we call this a dual page) at the same time. However, the recognition quality is higher if the dual page is split into two after scanning, so that each batch page corresponds to a single book page. Recognition and layout analysis are performed separately for each page, each page is deskewed if needed.

If you'd like to split the dual page:

- Select the **Split dual pages** option on the **Scan/Open Image** tab (menu **Tools>Options**) before scanning.

If you set this option, each dual page will be split into two batch pages. See "General Information on Working with Batches" section for more information about batches.

# Opening Images

If you don't have a scanner, you may still recognize image files (see the list of supported image formats in the "Supported Image Formats" section).

### To open an image:

- Click the arrow right to the **1-Scan** button and select the **Open Image** item in the local menu. The **1-Scan** button icon will change – the **Scan** caption will be replaced with the **Open** caption.

- Select the **Open Image** item in the **File** menu.

Select one or several images in the **Open** dialog. The selected images will be displayed in the **Batch** window and the last selected image will be displayed in the **Image** and **Zoom** windows of FineReader. All selected images are copied into the batch folder. See "General Information on Working with Batches" section for more details on how a batch is organized and how pages are displayed in it.

---

**Tip:** If you want the opened images to be recognized right away, use the **Open&Read** mode:
1. Select the **Open&Read** item in the **Process** menu or just press **CTRL+SHIFT+D.** The **Open** dialog will open.
2. Select the images for recognition in the **Open** dialog.

---

# Working with Image

- Despeckle image

- Invert image

- Rotate or flip image

- Clear block

- Increase/Decrease the image scale

- Get image information

- Undo the last action

**1. Despeckle image**

The recognized image may have lots of "dust" on it, i.e. it may contain lots of excess dots. These excess dots are due to scanning documents of medium or low print quality. Dots located close to character outlines may affect the recognition quality negatively. You may want to decrease the number of these dots. To do it:

- Select the **Despeckle Image** item in the **Image** menu.

If you want to despeckle a particular block:

- Select the **Despeckle Block** item in the **Image** menu.

---

**Note:** In case of a very faint original or an original set in a very light font, despeckling the image may cause periods, commas and very thin character parts to disappear, thus deteriorating the recognition quality.

---

If you scan or open "dusty" images, select the **Despeckle image** item in the

**Image preprocessing** group on the **Scan/Open Image** tab (menu
**Tools>Options**) to have images despeckled before the application adds
them to the batch.

**2.  Invert image**

Some scanners invert images (turning black color into white and vice versa)
during scanning.

You may want to apply the **Invert Image** option to make the document
appearance standard, i.e. black font against a white background. To do it:

- Select the **Invert Image** item in the **Image** menu.

---

**N o t e :** If you scan or open inverted images, select the **Invert image** item
in the **Image preprocessing** group on the **Scan/Open Image** tab (menu
**Tools>Options**) before adding these images to the batch.

---

**3.  Rotate or Flip image**

The recognition quality depends on the image having the standard orienta-
tion (the text should be read top-down and the lines should be horizontal).
By default FineReader detects the page orientation automatically during
recognition. If FineReader detected the page orientation incorrectly, clear
the **Detect image orientation (during recognition)** item on the **Scan/Open
Image** tab and **rotate the image** so that it has standard orientation:

- Click the ![] button or select the **Rotate Clockwise** item in the **Image**
  menu to rotate the image 90° clockwise.

- Click the ![] button or select the **Rotate Counter-Clockwise** item in
  the **Image** menu to rotate the image 90° counter-clockwise.

- Select the **Rotate Upside Down** item in the **Image** menu to rotate the
  image 180°.

To **flip** the image:

- horizontally (around the vertical axis) — select the **Flip Horizontal**
  item in the **Image** menu,

- vertically (around the horizontal axis) — select the **Flip Vertical**
  item in the **Image** menu.

**4.  Clear block**

If you don't want to recognize a certain image area or if you have large areas
of dust on the image, you may simply erase them. To do it:

- Select the  tool and then select the image area you want to erase holding down the left mouse button. Release the button to erase the selected image area.

**5. Increase/Decrease the image scale**

- Select the / tool on the **Image Tools** bar and click the image. The image scale will double/halve.

- Right-click the image and select the **Scale** item and then the scale you need in the local menu.

**6. Get image information**

You may get the following image information: image width and height in pixels; vertical and horizontal resolution per inch (dpi); image type.

- Right-click the image and select the **Properties** item in the local menu. A dialog will open. Select the **Image** tab in the dialog.

**7. Undo the last action**

- To undo the last action click the **Undo** button on the **Standard** bar .

---

**Tip:** To undo the **Undo** action click the **Redo** button on the **Standard** bar .

---

# Page Numbering

Each scanned page is given a number. The number given by default is the number of the last batch page plus one.

But you may set the number of the added page manually if you want to, for example, if you need to keep the original page numbering or when you need to scan a pile of numbered pages sorted by number. To do it:

- Select the **Ask for page number before adding page to the batch** item on the **Scan/Open Image** tab (menu **Tools>Options**) .

If you are scanning many double-sided pages sorted by number:

1. Select the **Ask for page number before adding page to the batch** item on the **Scan/Open Image** tab (**Tools>Options**).

2. Specify the number of the first scanned page in the **Page Number** dialog, then select the **Odd and even separately** option in the **Page numbering** field. Select the page numbering order: ascending or descending. The mode depends on the way you put your double-sided pages into the automatic document feeder, i.e. on whether the greatest or the lowest number is on top.

# Batch Image Options

Set these options on the **Scan/Open Image** tab (menu **Tools>Options**) before adding images to the batch.

**1.    Reduce color image resolution down to 100 dpi**

This option is used when you add big size color images to the batch. There are two advantages with this option: a) the batch will occupy less disk space, and b) image processing will take up less time.

---

**N o t e :** Setting this option doesn't affect the recognition quality. Only the quality of the saved pictures is affected, as their resolution is reduced down to 100 dpi.

---

**2.    Convert color and gray images to black and white**

Select the **Convert color and gray images to black and white** item in case you scan your images in grayscale using the **TWAIN-Source interface** and there are no color pictures, colored font and background on the scanned images, or you don't want to retain colors on the scanned images.
The scanned images will occupy less disk space if you select this option.

# Chapter 5

# Page Layout Analysis

ABBYY FineReader must "know" what image areas it needs to recognize before it starts recognition. It's the page layout analysis procedure that serves this purpose. This procedure draws text blocks, picture blocks, table blocks and barcode blocks (the latter is only available in the **Office** version).

In this chapter you will learn more about the following: when manual page analysis may be needed, what block types there are, how blocks drawn by automatic layout analysis procedure could be edited, and also how the layout analysis process could be made easier by using block templates.

## Chapter Contents:

- General information on page layout analysis
- Block types
- Automatic page layout analysis options
- Drawing and editing blocks manually
- Manual table layout analysis
- Using block templates

# General Information on Page Layout Analysis

Page layout analysis may be carried out both automatically and manually. In most cases, FineReader manages the complex task of page layout analysis itself. Start the automatic analysis by clicking the **2-Read** button. The recognition will be performed simultaneously with the layout analysis.

**Note:** A standalone procedure of page layout analysis is also available (menu **Process>Analyze Layout**). You may run this standalone procedure if needed, but remember that in this case the page layout analysis quality could be inferior, because the coupled layout analysis/recognition procedure run by the above-mentioned button uses additional information acquired during the recognition to aid layout analysis.

**You may want to draw blocks manually if:**

1. You want to recognize **only a part** of a page;

2. Automatic layout analysis drew blocks incorrectly.

**Tip:**
1. In some cases the quality of automatic analysis may be improved by adjusting the page layout analysis options. Check the current layout analysis options (**Recognition** tab, menu **Tools>Options**).
2. In case the application drew some blocks incorrectly, it is often faster to modify the incorrect blocks with the aid of the block editing tools instead of deleting all blocks and drawing them manually anew.

# Block Types

Blocks are image areas enclosed in frames. Blocks tell the system what image areas to recognize and in what order. The blocks also affect how the original page layout is retained. Blocks of different types have frames of different color. You can change the block frame colors on the **View** tab in the **Options** dialog (menu **Tools>Options**) in the **Appearance** group. Select the necessary block type in the **Item** field and the color you want in the **Color** field.

## The following block types are available:

**Recognition Area** – this block type is used for automatic recognition and analysis. After you click the **2-Read** button, all blocks of this type will be automatically analyzed and recognized.

**Text** – this block type is used for text image areas. It should only contain text formatted in one column. If there are pictures inside the text, draw separate blocks for them.

**Table** – this block type is used for table image areas or for areas of text structured into a table. When the application reads blocks of this type, it draws vertical and horizontal separators inside the block so as to form a table. This block is represented as a table in the output text. You can draw and edit tables manually.

**Picture** – this block type is used for image areas containing pictures. A block of this type may enclose an actual picture or any other object (say, a section of text) you want to be displayed as a picture in the recognized text.

**Barcode** (only available in the **Office** version) – this block type is used for barcode image areas. If your document has a barcode printed on it, and you don't want it to be displayed as a picture but as a series of letters and numbers in the recognized text, draw a separate block for the barcode and set the block type to **barcode**.

**N o t e :** There is an option enabling automatic barcode analysis and recognition, but it is not set by default. To set it, select the **Look for barcodes** item on the **Recognition** tab (menu **Tools>Options**).

# Automatic Page Layout Analysis Options

The FineReader automatic page layout analysis procedure draws blocks of the following types: text blocks, table blocks, picture blocks and barcode blocks (the latter is only available in the **Office** version).

Automatic layout analysis procedure (coupled with text recognition) is started when you click the **2-Read** button. Before clicking this button you should set the main layout analysis options: document type and table analysis options.

## Document type

In most cases the text layout is detected automatically. Automatic detection is performed if the **Autodetect layout** value on the **Recognition** tab in the **Document Type** group (menu **Tools>Options**) is set. This value is set by default.

In case you need to set the document type manually:

☛ Select the necessary item in the **Document type** group on the **Recognition** tab in the **Options** dialog (menu **Tools>Options**).

## Document types available:

**Autodetect layout** — it indicates that text layout is detected automatically. This value is set by default. If this value is set, recognition of all text types, including multi-column texts, texts with tables and pictures is performed automatically.

**Single column** — it indicates that the text is formatted in one column. Use this option if automatic page layout analysis incorrectly determined the text type as multi-column.

**Plain text formatted with spaces** — it indicates that that the text is formatted in one column and set in a monospaced font of a same size. In the recognized text left indents will be represented as spaces; every line is made a separate paragraph and the original paragraphs will be separated by empty lines. Use this type if, say, you recognize C++ code printouts or old computer printouts.

## Table analysis options

In most cases the application divides the table into rows and columns automatically. If you need additional tuning of table options, open the **Recognition** tab in the **Tables** group. Adjust these options if:

- the automatic page layout analysis procedure drew the table rows and columns incorrectly;

- the document contains many simple tables of a same type (e.g. there are no merged cells in the table or there is always only one line of text per cell).

1. Use the **One line of text per cell** option if you are dealing with a table without black separators or a table in which some separators are black whilst others are not, and which has only one line of text per every cell. An example:

| Kilometers | Miles |
|---|---|
| 1 | 0,62 |
| 5 | 3,1 |

— this table has only one line of text per cell

| Physical Phenomenon | t, degrees centigrade |
|---|---|
| Water boiling point | 100 |
| Water freezing point | 0 |

– this table has more than one line of text per cell
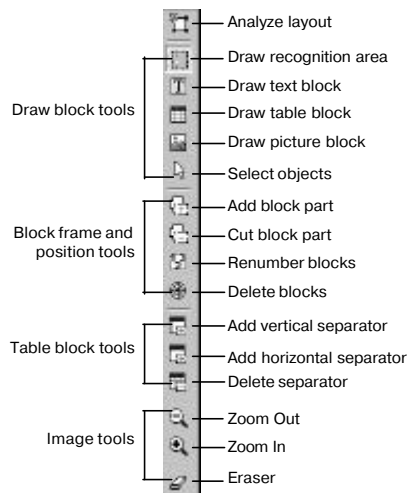
2. Use the **No merged cells in table** option if you are dealing with a table with no merged cells in it. An example:

| Temperature | |
|---|---|
| Degrees centigrade | Degrees Kelvin |
| -273 | 0 |
| 100 | 373 |

– the **Temperature** cell is a merged cell

---

**N o t e :** Do not select **One line of text per cell** and **No merged cells in table** options if there are tables of different structures in your text. Selecting these options may cause layout analysis mistakes and affect the recognition quality negatively.

---

# Drawing and Editing Blocks Manually

Draw block tools
- Analyze layout
- Draw recognition area
- Draw text block
- Draw table block
- Draw picture block
- Select objects

Block frame and position tools
- Add block part
- Cut block part
- Renumber blocks
- Delete blocks

Table block tools
- Add vertical separator
- Add horizontal separator
- Delete separator

Image tools
- Zoom Out
- Zoom In
- Eraser

## To create a new block:

1.  Select one of the following tools:

    ☒ – to draw a recognition area;

    ☒ – to draw a text block;

    ☒ – to draw a picture block;

    ☒ – to draw a table block.

2.  Position the mouse at the point where you want a corner of your block to be. Hold down the left mouse button and drag the mouse pointer to the point where you want the opposite block corner to be.

3.  Release the mouse button.

A frame will enclose the image area you selected.

You may then change the block type. The drawn block type may be one of the following: **Recognition Area**, **Text**, **Table**, **Picture**, or **Barcode**. To change the block type:

-   Right-click the block and select the **Block Type** item and then the corresponding block type in the local menu.

## Modifying Blocks

**To move the block borders:**

1.  Click the block border and hold down the left mouse button. The mouse pointer will become a two-headed arrow.

2.  Drag the pointer in the direction you need.

3.  Release the mouse button.

---

**Note:** If you click a block corner, you will drag both horizontal and vertical borders of the block at the same time.

---

**To add a rectangular block part:**

1.  Select the ☒ tool.

2.  Click the block you want to add a part to. Press and hold down the left mouse button then drag the mouse pointer diagonally. Select the image area you want to add to the block and release the button. The rectangle you've drawn will be added to the block.

3.  If necessary, move the block border.

**To cut out a rectangular block part:**

1. Select the ⊞ tool.

2. Click the block you want to cut a part from. Press and hold down the left mouse button then drag the mouse pointer diagonally. Select the image area you want to cut and release the button. The selected rectangle will be cut from the block.

3. If necessary, move the block border..

---

**N o t e :** If you cut a bottom or a top block part, you may only cut out a whole block corner. If you try to cut a part between the two upper or lower corners, the application will act as if you've cut the right block corner (respectively upper or lower). This behavior is due to the fact that the text lines must be unbroken or the recognition won't perform correctly. It is, however, no problem cutting side block parts.

---

**To select a block:**

- Select the ▱ tool and click the necessary block or drag the arrow over the blocks you want to select.

---

**N o t e :** You can select one or more blocks using the usual drawing tools. To select several blocks hold down **SHIFT** with one of the tools chosen:

▢, ⊤, ▣ or ▦ and drag the arrow over the blocks you want to select. To invert the selection (to add unselected block or to unmark the selected

one), hold down **CTRL** with one of the tools chosen: ▢, ⊤, ▣ or ▦ and drag the arrow over the the necessary blocks.

---

**To move blocks:**

- Hold down **ALT** with one of the tools chosen: ▱, ▢, ⊤, ▣ or ▦ and move the blocks.

**To renumber blocks:**

1. Select the ▯ tool.

2. Click the blocks in the order you want. The contents of blocks will be displayed in the output text in the same order.

**Note:** If you renumber blocks on an already recognized image, the recognized text in the **Text** window will be re-arranged accordingly.

**To delete a block:**

- Select the ⊞ tool and click the block you want to delete.

**Note:** If you delete an already recognized block, its text in the **Text** window will be deleted too.

**To delete all blocks on the image:**

- Select the **Delete Blocks and Text** item in the **Batch** menu.

**Note:** If the image you delete blocks on has already been recognized, all the recognized text in the **Text** window will also be deleted.

# Manual Table Layout Analysis

**Tip:** If the automatic table layout analysis procedure didn't draw the table rows and columns correctly, try to edit the automatic analysis results instead of deleting all blocks and drawing them manually anew. In most cases it will take up less time.

## To edit a table manually:

Use the following **Image** toolbar tools to edit the table:

⊞ Add vertical separator

⊞ Add horizontal separator

⊞ Remove separator

If the table cell contains only a picture, select the **Treat cell as a picture** item in the **Block Properties** dialog (menu **View>Properties**). If the table cell contains both text and pictures, draw a separate picture block (or blocks) inside the cell.

## To merge table cells or rows:

- Select the **Merge Table Cells** or **Merge Table Rows** item
  in the **Edit** menu.

**N o t e :**

1. You can split previously merged cells using the **Split Table Cells**
   command (**Edit** menu). The **Merge Table Rows** option doesn't affect
   the division of table into columns.
2. To avoid drawing horizontal and vertical separators manually, draw
   a separate table block, then right-click it and select the **Analyze Table
   Structure** item in the local menu. The system will draw all the separa-
   tors it considers necessary. You can then edit the table in case the sys-
   tem drew some separators incorrectly.

# Using Block Templates

If you are processing a bulk number of documents having the same layout,
like forms or questionaries, analyzing each page's layout separately will defi-
nitely take up a lot of time. To save your time you may create a block template,
a standard "set" of blocks of particular type representing the layout of your
pages, and then apply it to all the pages of the same layout you need to read.

## To create a block template:

1. Open an image and draw the blocks automatically or manually.

2. Select the **Save Blocks** item in the **Image** menu. The **Save Blocks as**
   dialog will open. Type a file name for the block template in the dialog.

## To load a block template:

1. Click the **Batch window** and select pages you want to apply
   the block template to.

2. Select the **Load Blocks** item in the **Image** menu. The **Open Blocks**
   dialog will open.

3. Select the block template file you need in the dialog.

4. Click the appropriate **Apply to** group radio button. The **All pages** radio
   button applies the block template to all the batch pages, the **Selected
   pages** radio button applies the block template to the selected pages only.

5. Click the **Open** button.

C h a p t e r   6

# Recognition

The aim of the OCR is to read text from the source image and to retain the source page layout as well. So, you should set the main recognition parameters - recognition language, the recognized text print type and document type - before you start recognition. This chapter deals with these parameters and also discusses other important recognition issues, illustrates the use of different recognition settings etc.

## Chapter contents:

- General information on recognition
- Recognition language
- Source text print type
- Other recognition options
- Background recognition mode
- Distributed document processing over a network
- Recognition with training
  - How to train a user pattern
  - How to edit a user pattern
- Creating a new language or a new language group
  - How to create a user language
  - How to create a new language group

# General Information on Recognition

**Note:** Check the following options: recognition language, the recognized text print type and document type before you start recognition.

**You may:**

1. Recognize a block or several blocks drawn on an image.

2. Recognize an open page or all pages selected in the **Batch** window.

3. Recognize all unrecognized batch pages.

4. Recognize all the pages in the background mode. This mode lets you edit and recognize pages at the same time.

5. Recognize pages in the train mode. This mode is usually used for recognizing texts set in decorative fonts or for processing large volumes (more than a hundred pages) of documents of inferior print quality.

6. Recognize the same batch on several workstations.

**To start recognition:**

- Either click the **2-Read** button on the **WizardBar** toolbar, or

- Select the item you need in the **Process** menu:

  **Read** – to recognize the open page or all the pages selected in the Batch window;

  **Read All Page**s – to recognize all unrecognized batch pages;

  **Read Block** – to recognize a block or several blocks drawn on the image;

  **Start Background Recognition** – to start recognition in the background mode.

By default, the **2-Read** button recognizes the open image. To change the button mode, click the arrow right to it and select the necessary item in the local menu.

**Note:** When you perform OCR on a block that had been recognized already, recognition will only be applied on the modified and added blocks.

# Recognition Language

FineReader recognizes both mono- and multi-lingual (e.g. English-French) documents.

To set the text recognition language, select it in the drop-down list on the **Standard** toolbar.

If you want to recognize a multi-language document:

1. Select the **Select multiple languages** item in the language list on the **Standard** toolbar. The **Recognition language** dialog will open.

2. Select the necessary languages in the **Recognition language** dialog.

---

**N o t e :**
1. Increasing the number of the recognition languages used simultaneously may affect the recognition quality negatively. The reasonable number of languages to be used simultaneously is 2-3.
2. (for users of the FineReader **Office** version only) If you often use a certain language combination, you can create a new language group and add the languages you need to it.
3. Check the fonts selected on the **Formatting** tab before you read your documents: these fonts must support all characters used in the recognition languages you select, otherwise the recognized text will be displayed incorrectly (there will be "?" or "□" symbols instead of letters).

---

Sometimes you won't find the recognition language you need in the language list. It may be due to the following reasons:

1. The language is not supported by FineReader. See the full list of the recognition languages in the "Supported Languages" section of ABBYY FineReader Help.

2. The language isn't included in the recognition language list displayed on the **Recognition** toolbar. To add the language you need select the **Choose more languages** item in the language list on **Standard** toolbar. The **Recognition language** dialog will open. Select the language you need in the dialog.

3. The language has never been installed (the minimal installation mode was used) or the language had been disconnected during the custom installation. To install additional languages run the FineReader installation program in the component installation mode (**Custom installation**), clear all the items except **Recognition languages** and click the **Details** button. A list of languages supported by FineReader will be displayed. Select the languages you need.

**Note:** Make sure you've indicated the same folder you've installed FineReader in.

**To show/hide a language in the drop-down list on the toolbar:**

- Select the language you need in the **Language editor** dialog (menu **Tools>Language Editor**) and then select or clear the **Show this language in the drop-down list on the toolbar** item.

**Tip:** You may even set a recognition language for a separate block! To do it, right-click the block you want to set a different recognition language for and select the **Properties** item in the local menu. The **Properties** dialog will open. Select the **Block** tab in the dialog and then select the block recognition language in the **Languages** field on the tab.

# Source Text Print Type

In most cases the text print type is detected automatically. To use the automatic print type detection select **Autodetect** in the **Print Type** group (menu **Tools>Options, Recognition** tab).

If you want to recognize draft mode dot matrix printer printouts or typewriter printouts, you may increase the recognition quality by setting another **print type**:

- Select the **Typewriter** item if you want to recognize typewriter printouts,
- Select the **Dot Matrix Printer** item if you want to recognize dot matrix printer printouts.

software — An example of draft mode dot matrix printer printout. You see that the character lines are made up of separate dots.

software — An example of a typewriter printout. All the letters are of equal width (compare, for example, "w" and "a").

**To change the print type:**

- Select the item you need on the **Recognition** tab in the **Options** dialog (menu **Tools>Options**).

**Note:** After you are done recognizing typewriter or dot matrix printer printouts, do not forget to set the **Autodetect** item back again to recognize normal texts correctly.

# Other Recognition Options

**Show image during recognition**

Recognition of a large number of pages is faster if the processed image is not displayed. To run recognition without displaying the image:

- Clear the **Show image during recognition** item on the **General** tab (menu **Tools>Options**).

**Invert block**

To read inverted (white font against black background) image areas – a text block, a table cell or a whole table:

- Right-click the necessary block and select the **Properties** item in the local menu. The **Block Properties** dialog will open. Select the **Inverted** item in the **Block Properties** dialog.

# Background Recognition Mode

If you want to edit the already recognized pages and run the recognition at the same time, you may find the background recognition mode useful. To start background recognition:

- Select the **Start Background Recognition** item in the **Process** menu.

  The 🖳 sign will appear in the status line at the bottom of FineReader main window. If the **Batch** window has **Details** view mode on (turn the **Details** view mode on by right-clicking the **Batch** window and selecting **View>Details** item in the local menu), the page currently being recognized will have the 🖳 icon displayed in the **Opened by** column.

When the background recognition mode is on, recognition will resume automatically if an unrecognized page is added to the batch.

To stop **Background Recognition:**

- Select the **Stop Background Recognition** item in the **Process** menu.

---

**N o t e :** The background recognition mode uses the currently set recognition options.

---

# Distributed Document Processing over a Network

**(only available in ABBYY FineReader Office version)**

For inputting input bulk volumes of documents, an important factor is the throughput of the recognition system you use. FineReader enables you to increase the throughput nearly indefinitely thanks to its support of the distributed document processing over a network.

The document input process consists of the following steps: scanning, recognition, editing and saving. Each step or any number of these steps may be performed on any number of workstations. For example, a workstation with an ADF scanner may be assigned the single task of scanning documents. At the same time another workstation may be assigned the recognition, editing and/or saving of the read pages.

The same batch may be processed simultaneously by two or more workstations, so that the whole process speeds up quite considerably. A copy of FineReader must be installed on each workstation processing the batch (see "Network Installation" section for more details on FineReader network installation) and each workstation processing the batch should have a read-write access to the batch folder.

You may view each page's status (e.g. whether the page has been scanned, recognized, edited and/or saved, which network user performed the editing, scanning etc.) in the **Batch** window. To turn on the **Batch** window's **Details** mode, right-click the **Batch** window and select the **View>Details** item in the local menu.

If there are several workstations processing a batch, the system will automatically distribute the work among them. The process works like this. First the scanning workstation adds an unrecognized page to the batch. Then, if there is a "free" recognition workstation working in background recognition mode at the same time, it immediately begins recognizing this page and puts a special mark in the batch, so that this page becomes "locked" for all the other recognition workstations. If all recognition workstations are busy at this time, the scanned page remains unnoticed until a recognition workstation finishes recognizing a page and starts looking for new pages.

At the same time any other workstation may also open already recognized pages for checking, editing and saving.

**How to process a batch over a network:**

1. Create/open a batch and set the following batch options: recognition language, source text print type, page layout, table analysis options, and, if necessary, the user pattern.

2.  Start FineReader on all the workstations you want to recognize the batch on, and then open the batch on each of them. Start background recognition (menu **Process>Start Background Recognition**) on all of them.

3.  Start scanning on the workstation with an ADF scanner.

---

**Tip:** If your scanner doesn't support TWAIN (that's often the case with high-speed scanners), you can scan directly into the FineReader batch folder. To do it, the workstation with an ADF scanner should scan the images using any image acquiring application the scanner supports, then it should save the scanned images into the batch folder. Scanned file names must look like this: 0001.tif, 0002.tif, 0003.tif etc. The numbers should reflect the order in which you scanned your pages.

---

FineReader will automatically "pick up" all the scanned images and recognize them. Press **F5** or select the **Update Page List** item in the **Batch** menu to have the page list updated.

4.  You may edit the recognized text and save it to a file or an application you need on the workstations running the background recognition mode or on any other workstation.

---

**Note:** If you choose to save in **TXT, CSV** or **DBF** formats, you may append the recognized text to an already existing file. To do it, use the **Append to the end of file** option (see TXT, CSV and DBF tabs in the **Formats Settings** dialog).

---

# Recognition with Training

As mentioned earlier, FineReader has no problem reading texts set in practically any font and is nearly insensitive to printing quality, so usually you may recognize any text without prior training. However, FineReader does provide you with user pattern training tools.

**You may want to use the Train User Pattern mode for:**

1.  recognizing texts set in decorative fonts;

2.  recognizing texts containing special characters (e.g. mathematical symbols);

3.  recognizing large volumes (more than a hundred pages) of texts of low print quality.

**Tip:** Use the **Train User Pattern** mode **only in the above-stated cases.** In other cases you may get a tiny increase in the recognition quality, but it won't be worth the time spent on pattern training.

The pattern training process works like this. Recognize a page or two in the special training mode. The result of this training is a pattern. FineReader will then use this pattern as an aid during recognition of the rest of the text.

Sometimes two or even three characters may get "glued" together. In this case it is often impossible to separate them – FineReader cannot enclose them in frames so that each frame contains only one character. If this is the case (i.e. you cannot move the frame so that it contains an entire character and at the same time does not contain parts of other characters next to it), you may train FineReader to recognize the whole "inseparable" character combination. Examples of typical inseparable combinations are ff, fi, fl etc. Such combinations are called ligatures.

**Note:**
1. The created pattern is useful only for recognizing texts set in the same font of the same size and scanned with the same resolution as the document you trained your pattern on.
2. Each pattern is created for a particular batch, so if you delete a batch the pattern is deleted as well. You may make any pattern available in other batches. To do it, save the batch options in a file of batch template format file type.
3. If you switch to recognizing texts set in a different font, don't forget to disconnect the pattern - click the **Do not use user pattern** radio button on the **Recognition** tab, menu **Tools>Options.**

**If you want to train a user pattern:**

1. Start the **Train User Pattern** mode - click the **Train user pattern** radio button on the **Recognition** tab, menu **Tools>Options**, in the **Training** group. The default pattern name ("Default") will be displayed in the status bar.

2. Click the **2-Read** button.

3. Train your pattern - recognize a page or two in the **Train user pattern** mode. Trained characters are saved in the default pattern. After you are finished training the pattern, FineReader will save the created pattern (**Default.pat**) in the current batch folder.

4. Edit your pattern.

5. Turn the train mode off (click the **Use user pattern** radio button on the **Recognition** tab).

6. Recognize the rest of the text – click the **2-Read** button.

**Note:**

1. To create several patterns for the same batch use the **Pattern Editor** dialog (click the **Pattern Editor** button on the **Recognition** tab or select the **Tools>Pattern Editor** menu item). Create a new pattern (click the **New** button in the dialog) and then select it (click the **Set Active** button). Working with a created pattern is the same as working with the default pattern, see steps 1-5. Keep in mind that only one pattern may be active at the same time.

2. If you've created several patterns for the same batch, the active one is the last created pattern. The active pattern name is displayed in the status bar. To activate another pattern, select the pattern you need in the pattern list in the **Pattern editor** dialog (menu **Tools>Pattern Editor**) and click the **Set Active** button. Then click the **Use user pattern** radio button on the **Recognition** tab, menu **Tools>Options**, in the **Training** group.

3. If the **Use built-in patterns** option is set, FineReader tries to read the text using the built-in patterns and stops only at uncertain characters. If you are training the system to read decorative and/or non-standard fonts (for example, Tibetan) FineReader may read characters incorrectly using the built-in patterns. In this case clear the **Use built-in paterns** checkbox on the **Recognition** tab and train the system to recognize each character.

# How to Train a User Pattern

1. Make sure the **Train user pattern** radio button on the **Recognition** tab (menu **Tools>Options**) in the **Training** group is on.

2. Click the **2-Read** button. FineReader will start recognition. When it comes across a character it doesn't know, the **Pattern Training** dialog will open. The character image will be displayed in it.
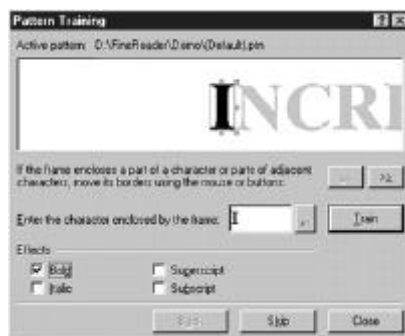
## Training to recognize a character:

The frame in the top dialog window should enclose **a single character**, and this character **must be fully enclosed** by the frame. If the frame encloses only a character part or more than one character, click the frame borders

and move them so that the above-stated requirements are met. The ⧸

and ⧸ buttons move the frame border as well (and are useful for training italic symbols, see below). After you are done, type the character and click the **Train** button.

---

### N o t e :

1. You may only train the system to read the characters included in the language alphabet.
   If you train FineReader to read characters you cannot enter from the keyboard, you may use a combination of two characters to denote these "unenterable" characters or you may copy the necessary

   character from the **Character Table** (click the ⧸ button in the **Pattern Training** dialog to open it).

2. If you'd like to train the system to retain character formatting, select the corresponding **Italic** or **Bold** item in the **Pattern Training** dialog before clicking the **Train** button.

3. Make sure that you type uppercase characters when you train capital character images, and lowercase characters when you train lowercase character images.

---

If you have made a mistake during training, you may click the **Back** button and the frame will go back to the previous position, and the latest pair "image-character" will be removed from the pattern. This "undo" function is limited: you may only "undo" training in the last word.

## Training to recognize ligatures

A ligature is a combination of two or three "glued" characters, for example, fi, fl, ffi etc. These characters are difficult to divide into separate ones because they are "glued" in print. They are better treated as a "single" compound character.

Training ligatures is the same as training separate characters:

1. Type the necessary character combination and click the **Train** button.

2. The frame in the top dialog window should enclose **the entire ligature**. You can move the frame border using the mouse or by clicking

   the ⧸ and ⧸ buttons.

Each pattern may contain up to 1000 new characters. However, you should not create too many ligatures, as it may affect the recognition quality negatively.

**You should also take the following limitations into account when you train ABBYY FineReader:**

1. FineReader doesn't differentiate between certain characters which are usually considered different. Such images are recognized as one and the same character. For example, the straight ('), right (') and left (') apostrophes are kept in the pattern as one character – the straight apostrophe. Thus, you'll never see the right and left apostrophes in the recognized text, even if you try to train them.

2. In some cases a certain image is recognized as a certain character depending on its environment.

# How to Edit a User Pattern

You may find useful to edit your newly created pattern before you start recognition with it. An incorrectly trained pattern will affect the recognition quality negatively, so it is useful to check your pattern.

The pattern should contain only whole characters or ligatures. Characters with cut edges and incorrectly labelled characters should be removed from the pattern.

**To edit a user pattern:**

1. Select the **Pattern Editor** item in the **Tools** menu. The **Pattern Editor** dialog will open.

2. Select the necessary pattern and click the **Edit** button in the dialog. The **User Pattern** dialog will open.

3. Select a character and click the **Properties** button to edit the character caption and set the correct typeface: italic, bold, subscript or superscript. You may also click the **Delete** button to remove incorrectly trained characters from the batch.

# User Languages and Language Groups

**(only available in the ABBYY FineReader Office version)**

In addition to the built-in languages and language groups, you may also create a new language or make up a new language group (from languages supported by FineReader) and use it for recognition.

**You may want to create a new language if you need:**

1. To connect a user dictionary.

    - For example, you need to recognize an English text containing many abbreviations. You may create an abbreviation dictionary, create a new language and connect the dictionary to the language. Then you may make up a new language group consisting of the English language (with the system dictionary) and your new language (with the abbreviations dictionary) and use this language group to recognize your texts.

2. To recognize documents of certain special types, for example:

    - supermarket product lists where there are only product codes. Such codes usually contain only numbers and probably several letters. You may then create a new language consisting of the necessary characters only and use it for recognizing these documents.

    - documents set in capitals only. You may increase the recognition quality if you create a language in which all lower-case letters are prohibited.

You should create a language group if you use a particular language combination often.

To create a new language or a language group open the **Language Editor** dialog (**Tools** menu, **Language Editor** item).

# How to Create a User Language

**(available in ABBYY FineReader Office version only)**

**To create a new recognition language:**

1. Select the **Language Editor** item in the **Tools** menu.

2. Click the **New** button and in the opened dialog select the **Create a copy of the language** radio button, then select a source language for the new one.

3. The **Simple Language Properties** dialog will open.
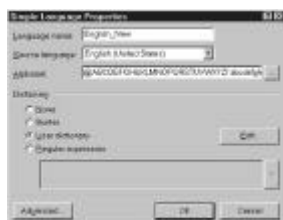
**Set the following new language parameters (all the parameters are set in the Simple Language Properties dialog):**

1. The new language name.

2. The basic alphabet to be used by your new language. This parameter is set in the **Alphabet** field. If necessary, edit the alphabet by clicking the ⊡ button.

3. The dictionary to be used by the system (for both recognition and spelling check). You may choose one of the following alternatives:

   - **None** (no dictionary for the language).

   - **Built-in** (a dictionary supplied with FineReader)

   - **User dictionary**

     To add words to the dictionary or to connect an old user dictionary or a text file in Windows (ANSI) or Unicode encoding (the only requirement is that the words must be separated by spaces or other non-alphabetic characters) click the **Edit Dictionary** button.

     ---

     **Note:** The spelling check considers the user dictionary words correct if they are found in the text in one of the following capitalizations: dictionary set capitalization, all letters small, all capitals, the first letter is capital, the rest small. Some examples are:

     | **Dictionary set capitalization:** | **Correct occurrences of the word:** |
     |---|---|
     | abc | abc, Abc, ABC |
     | Abc | abc, Abc, ABC |
     | ABC | abc, Abc, ABC |
     | aBc | aBc, abc, Abc, ABC |

     ---

   - **Regular expression** (specifies the new language grammar; see the **Regular Expressions** section of ABBYY FineReader Help for details).

---

**Note:** You may click the **Advanced** button in the **Simple Language Properties** dialog to set some advanced properties of the new language, for example, characters ignored if in words, prohibited characters, etc.

---

# How to Create a New Language Group

**(only available in the ABBYY FineReader Office version)**

If you often recognize texts written in a certain language combination, say, English-German, you may create a language group combining these languages. The created group will be displayed in the language list on the **Standard** toolbar.

**N o t e :** You may specify the recognition languages to be used in the language list on the **Standard** toolbar. To do it, select the **Select multiple languages** item in the list. The **Recognition language** dialog will open. Select the languages you need in the dialog.

## To create a recognition language group:

1.  Select the **Language Editor** item in the **Tools** menu and click the **New** button. A dialog will open. Select the **Create a new group of languages** item in the dialog.

2.  The **Language Group Properties** dialog will open.



**Set the following new language group parameters (all the parameters are set in the Language Group Properties dialog):**

1.  Group name.
2.  Languages making up the group.

**N o t e :** In some cases you are sure that certain characters will never be found in the texts you recognize. You may then specify the set of such characters (called prohibited characters) in the language group properties. Specifying such characters may increase the speed and quality of the recognition. To specify the set of prohibited characters, click the **Advanced** button in the **Language Group Properties** dialog. The **Advanced Language Group Properties** dialog will open. Specify the set of prohibited characters in the **Prohibited characters** line.

Chapter 7
# Text Checking and Editing

After recognition is over, you will see the recognized text in the **Text** window. The **Text** window is ABBYY FineReader's built-in editor, here you can check the recognition results and edit the recognized text.

### The FineReader text editor has two distinctive features:

1. A built-in spelling check system (see the list of languages with spelling check support in the "Supported Languages" section of ABBYY FineReader Help).

2. A convenient visual aid: the source image of the text line you edit is displayed in the **Zoom** window.

### The built-in spelling check system features:

1. Tools for finding uncertain words (words with uncertain characters).

2. Tools for finding misspelt words.

3. Tools for adding unknown words to the ABBYY FineReader dictionary. Adding words improves the recognition quality.

### Chapter contents:

- Checking text in ABBYY FineReader
- Check and edit text options
- Adding and deleting words from the user dictionary
- Editing text in ABBYY FineReader
- Editing tables

# Checking Text in ABBYY FineReader

Uncertainly recognized characters and words not found in dictionary are highlighted with different colors. By default, the light blue color is used for uncertain characters and pink — for words not found in the dictionary. To change these colors:

1. Select the **Uncertain Character** (or **Not in Dictionary word**) item and the color you want in the **Color** item on the **View** tab (menu **Tools>Options**) in the **Appearance** group.

**To check the recognition results:**

1. Click the **3-Check Spelling** button on the **WizardBar** toolbar (or select the **Check Spelling** item in the **Tools** menu).

2. The **Check Spelling** dialog will open.
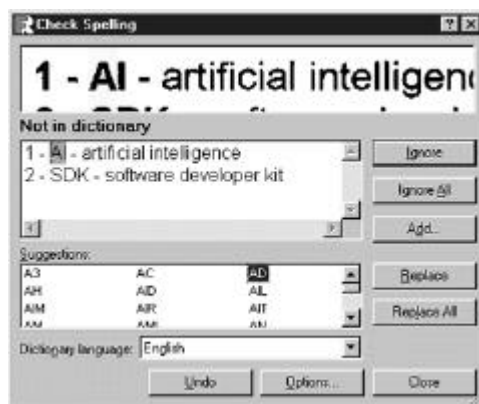


3. There are three windows in the **Check Spelling** dialog. The top window is the same as the FineReader **Zoom** window, it displays the original image of the possibly misspelt word. The middle window displays the possibly mis-spelt word itself, whilst the line above it displays the name of the possible misprint type. The **Suggestions** window at the bottom provides you with replacement suggestions (if there are any). The suggestions are based on the dictionary selected in the **Dictionary language** drop-down list; you may select any language you want from this list.

**N o t e :** You can enlarge the **Check Spelling** dialog for your checking and editing convenience. To do it, simply click the dialog border; the mouse pointer will become a double-headed arrow. Drag the border to make the dialog larger or smaller.

4. What can you do with the possibly misspelt word?
   Here are the alternatives:
   - Click the **Ignore** button to leave the word unchanged.

   - Click the **Ignore All** button to leave all such words in the text unchanged.

   - Select a replacement suggestion and then click the **Replace** or **Replace All** button to replace the current word or all such words in the text. If there is no correct suggestion for the word in the **Suggestions** window, type the correct one in the middle dialog window and then click the **Confirm** button to change the current word.

   - Click **Add…** to add the word to dictionary. After the word is added, the application will consider all subsequent occurrences of this word in any of its word forms to be correct.

   - Click **Options…** to set the spelling check options.

   - Click **Close** to close the dialog window.

# Check and Edit Text Options

These options are set on the **Check Spelling** tab (menu **Tools>Options**).

1. Error display level

   **N o t e :** This option must be set before you start recognition.

2. Stop at words with uncertain characters

3. Stop at words not found in dictionary

4. Stop at compound words

5. Ignore words with digits and other nonalphabetic characters

6. Correct spaces before and after punctuation marks

## Error display level

In the **Error display level** list you may select the following values:

1. **None** – recognition errors are not highlighted.

2. **Standard** – unrecognized, and characters uncertainly recognized are highlighted.

3. **Thorough** – the same as **Standard**, not found in the dictionary words are highlighted too.

**Note:** The number of errors displayed in the **Text** window will change after you re-read the page.

### Stop at words with uncertain characters

If you set this option, the spelling check will stop at words with uncertain characters.

### Stop at words not found in the dictionary

If you set this option, the spelling check will stop at words not found in the dictionary. If a word is not found in the dictionary, then it may have been read incorrectly.

### Stop at compound words

If you set this option, the spelling check will stop at those words not in dictionary which can be either made up according to the available morphology models or from the words in the dictionary.

### Ignore words with digits and other nonalphabetic characters

If you set this option, the spelling check will treat all words with digits and other not included in recognition language characters as correct, unless there are uncertain characters in them.

### Correct spaces before and after punctuation marks

If you set this option, the spelling check won't stop if it finds incorrect spaces before or after punctuation marks, it will just correct the spaces automatically.

# Adding and Deleting Words from the User Dictionary

### Adding a word to the user dictionary

Enlarging the dictionary is a good way to increase the recognition quality. During recognition, FineReader checks the words it recognizes against the dictionary. Thus it is good idea to add the most frequent unknown words (e.g. terms, abbreviations, names etc. frequently used in your documents) to the user dictionary.

A distinctive feature of FineReader's spelling check system is that it doesn't just add the word in the form it is first found in the text — it creates the word's paradigm, i.e. the set of all its forms. This feature enables FineReader to "know" the word in its entirety immediately after its first word form is found in the text, so it recognizes the word in all subsequently found forms.

**To add a word to the dictionary during the spelling check:**

☞ Click the **Add** button in the **Check Spelling** dialog.

Set the following parameters in the **Primary Form** dialog:

1. **Part of speech** (Noun, Adjective, Verb, Uninflected).

2. If the word is always spelt with an initial capital letter, select the **Proper name** item. If you add an abbreviation, select the **Abbreviation** item.

3. The primary form of the word.

Click **OK.** The **Create Paradigm** dialog will open. FineReader asks you various questions about the word forms in this dialog. Your answers will help FineReader to build the paradigm of the word you add. Click either **Yes** or **No** when FineReader asks you a question. If you've made a mistake, click the **Anew** button to answer this question correctly. The built paradigm will be displayed in the **Paradigm** dialog.

**N o t e :**
1. If you don't want to build the paradigm and always add always want to add words as uninflected, then select the **Add without prompting for word forms** option (English dictionary only) on the **Check Spelling** tab (menu **Tools>Options**).
2. You may also add words when you view the added words list. To do it, select the **View Dictionaries** item in the **Tools** menu. The **Select Language** dialog will open. Select the language you need in the **Select Language** dialog and click **View**. The dictionary with the list of the added words will open. You may add your word by clicking the **Add** button.

If the added word already exists in the dictionary, the system will warn you about it. In this case you can view its paradigm. If you think the existing paradigm is incorrect (which is often the case when you add a homonymous word belonging to another part of speech), you may build another one (click the **Add** button in the **Add Word** dialog).

**T i p :**
1. FineReader lets you import user dictionaries created by older versions (FineReader 3.0 and 4.0).

2. FineReader also lets you import your user dictionary (*.dic) created by Microsoft Word 6.0, 7.0, 97 and 2000.

**To import a dictionary:**

1. Select the **View Dictionaries** item in the **Tools** menu, then select the dictionary language and click the **View** button.

2. Click the **Import** button in the **View Dictionaries** dialog and select files with *.ext, *.txt or *.dic extensions.

**To delete a word from the dictionary:**

1. Select the **View Dictionaries** item in the **Tools** menu. Select the language you need and click the **OK** button. A dialog will open.

2. Select the word you want to delete and click the **Delete** button.

# Editing Text in ABBYY FineReader

**N o t e :** If the FineReader **Text** window doesn't display characters correctly (i.e. if there are "?" or "□" signs instead of all or some letters), it means that your current font doesn't support your recognition language alphabet in full. Select a font that supports your entire recognition set (for example, Arial Unicode or Bitstream Cyberbit) on the **Formatting** tab (menu **Tools>Options**) in the **Fonts** group and recognize the document again.
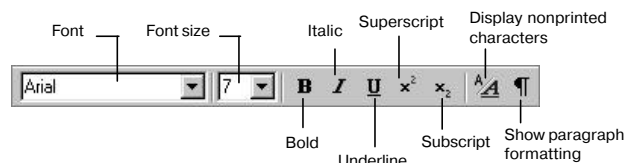


After a page is read, its text is displayed in the **Text** window.

**N o t e :** FineReader retains the original text formatting but doesn't display it in full in its text editor. The FineReader text editor displays the text as follows: arrangement into text lines is retained; the recognized text, pictures and tables are arranged in the order their source image blocks are numbered; sections read from different blocks are separated by dotted lines. When you send your text to an external application, the text layout is retained according to the layout retention options. Set these options on the **Formatting** tab (menu **Tools>Options**) and in the dialogs of respective formats.

Uncertainly recognized characters are highlighted. You may cancel high-lighting of such characters, just unselect the **Highlight uncertain characters** item on the **View** tab (menu **Tools>Options**).

If you'd like to make viewing the document in the **Text** window even more convenient, you may switch to the draft editor mode. The following text features aren't displayed in the draft mode: pictures (you see the "picture block" words in place of pictures); left indent; paragraph alignment (all paragraphs are aligned left); text and background color. A font of the same size (12pt by default) is used throughout to display the text in draft mode. The font and its effects (bold, italic, underlined, superscript and subscript) are retained in the draft mode. Switch between the draft and full modes by clicking the  button on the **Formatting** bar. If the button is pressed, the full mode is on, if not − the draft mode is on.

You may change the font size for displaying the text in draft mode.
To do it:

1. Select the **Options** item in the **Tools** menu.

2. Set your preferred font size in the **Draft editor font size** item on the **View** tab.

**The FineReader built-in editor provides the following text editing features:**

- Copy, cut and paste

- Search and replace

- Font effects

- Undo and redo

## Copy, cut, paste

1. Before you use the copy, cut and paste commands please select the necessary text.

2. Follow the instructions below depending on what you want to do:

   **To copy the selection:**

   - Either click the **Copy** button  on the **Standard** toolbar, or

   - Select the **Copy** command in the **Edit** menu or in the local menu, or

   - Press **CTRL+C**

**To cut the selection:**

- Either click the **Cut** button ✄ on the **Standard** toolbar, or

- Select the **Cut** command in the **Edit** menu or in the local menu, or

- Press **CTRL+X**

**To paste the copied text:**

- Either click the **Paste** button 🗐 on the **Standard** toolbar, or

- Select the **Paste** command in the **Edit** menu or in the local menu, or

- Press **CTRL+V**

## Search and replace

**To find a word or a phrase in the text you are editing:**

1. Perform one of the following actions:

    - Either select the **Find** item in the **Edit** menu, or

    - Press **CTRL+F**

2. The **Search** dialog will open. Type the word or the phrase you want to find in the **Find what** line of the dialog and set the search parameters.

---

**Note:** To search for the same word with the same parameters just press **F3**.

---

**To search and replace a word or a phrase in the text you edit:**

1. Perform one of the following actions:

    - Either select the **Replace** item on the **Edit** menu, or

    - Press **CTRL+H**

2. The **Replace** dialog will open. Type the word or the phrase you want to find in the **Find what** line of the dialog, type the word or the phrase you want to replace the search pattern with in the **Replace with** line, and set the search parameters.

## Font effects

1. Click the word or select the text you want to change the font for.

2. Perform one of the following actions:

    - Either click the button (e.g. **B**) you need on the **Formatting** bar, or

- Right-click the **Text** window and select the **Character Properties** item in the local menu. The **Character** dialog will open. Select the font type you need and set the necessary font parameters in the dialog, or

- Press **CTRL+B** – to apply bold, **CTRL+I** – to apply italics, **CTRL+U** – to underline the word or text.

**N o t e :** You can also set the following additional text formatting parameters in the **Font** dialog: character spacing, character scale and use of small caps. But keep in mind that FineReader built-in editor doesn't *display* such formatting; you will only see these changes after you save your document to an application supporting these formatting options (for example, MS Word).

## Undo and redo

Perform one of the following actions:

**To undo an action:**

- Either click the **Undo** button on the **Standard** toolbar, or

- Select the **Undo** item in the **Edit** menu, or

- Press **CTRL+Z**

**To redo an undone action:**

- Either click the **Redo** button on the **Standard** toolbar, or

- Select the **Redo** item in the **Edit** menu, or

- Press **CTRL+Y**

# Editing Tables

### The table editor provides you with tools to do the following:

- Merge cell or row contents

- Split cell contents

- Split row or column contents

- Delete cell contents

### To merge the cell or row contents:

- Holding down the **CTRL** button select the cells or rows you want to merge, then select the **Merge Table Cells** or **Merge Table Rows** item in the **Edit** menu.

### To split cell contents:

- Select the **Split Tale Cells** item in the **Edit** menu.

**N o t e :** This command may only be applied to cells merged previously.

### To split row or column contents:

1. Select the 🖳 or 🖳 tool on the toolbar in the **Image** window, then click the row/column you want to split to add a new horizontal/vertical separator to.

**T i p :** You can merge row contents by using the tool 🖳 or the **Merge Table Rows** command (**Edit** menu).

### To delete cell contents:

- Select one or several cells you want to delete in the **Text** window and press **DEL**.

# Saving into External Applications and Formats

The recognition results can be saved to a file, sent to an external application without saving, copied to the clipboard or sent by e-mail. You may either save all the pages or the selected ones only.

## ABBYY FineReader exports the recognition results into the following applications:

Microsoft Word 6.0, 7.0, 97 and 2000; Microsoft Excel 6.0, 7.0, 97 and 2000; Corel WordPerfect 7.0, 8.0 and 9.0;  Lotus Word Pro 9.5  and 97; StarWriter 4.x and 5.x, PROMT 98 and any other application supporting the ODMA standard.

## Chapter contents:

- General information on saving the recognized text
- Text saving options
- Saving the recognized text in RTF and DOC formats
- Saving the recognized text in PDF format
- Saving the recognized text in HTML format
- Saving the page image

# General Information on Saving the Recognized Text

### You may:

- save the recognized text using the **Save Wizard**,
- save the opened or selected pages to a file or send them to an external application,
- save all the batch pages to a file or export them into an external application,
- save the page image.

Click the **4-Save** button to send the recognition results to the application you need or save them to a file. The icon appearance depends on the saving mode currently on. The **Save** button caption displays the name of the currently selected export application.

**To save the recognized text:**

- Click the arrow right to the **4-Save** button and select the item you need in the local menu.

**N o t e :** If you'd like to save only a certain number of pages, select them before clicking the **4-Save** button.

After the export is done, the **4-Save** button icon will change depending on what you've done – sent the recognized text to an application, sent it by e-mail, copied it to the clipboard or saved it to a file. The **4-Save** button icon always adopts the icon of the last export mode you used. So if you'd like to export another image(s) using the same mode just click the icon itself and don't bother browsing the button's local menu.

# Text Saving Options

Text saving options are set on the **Formatting** tab in the **Tools>Options** menu. Some saving options may be set in the **Save Wizard** and **Save Text as** dialogs as well.

- Formatting and text layout retention modes
- Retain pictures
- Fonts to use
- Save all batch pages or selected ones only
- Recognized text saving modes

## Formatting and text layout retention modes (saving in RTF, DOC or HTML formats)

- **Retain full page layout** – the document layout is retained in full: arrangement into paragraphs, font and font size, columns, text direction, text color, the structure of tables.

- **Retain font and font size** – structure of tables, arrangement into paragraphs, font and font size are retained.

- **Remove all formatting** – only structure of tables and arrangement into paragraphs are retained.

---

**N o t e :** You may set some advanced options depending on the export format you choose. For example, in case of **RTF/DOC** formats you may set the default page size and uncertain character highlight mode; in case of **HTML** format you may set the picture resolution and the code page; in case of **PDF** format you may set the Type 1 fonts to use (if you use a code page different from the Latin one, e.g. Cyrillic, Greek etc.). You may set these options in the **Formats Settings** dialog (menu **Tools>Formats Settings**). The dialog has a separate tab for each format, just click the format tab you need and set the options you like.

---

## Retain pictures

If you set this option, the pictures will be saved together with the recognized text. The option is available if you save in **RTF, DOC** or **HTML** formats.

## Fonts to use (available if you save in RTF, DOC or HTML formats)

By default the fonts from the **Formatting** tab are used when you save in RTF, DOC or HTML format. You may specify what fonts to use. You may change the fonts in the **Text** window or you may set other fonts on the **Formatting** tab in the **Fonts** group and re-read the document.

## Save all batch pages or the selected ones only

You may either save all the batch pages or only the ones you select. If you'd only like to save a certain number of pages, select them before saving.

## Recognized text saving modes

**(available if you save several batch pages at a time)**

- **Create a separate file for each page** – each batch page is saved to

a separate file. The batch page number is automatically added to the end of the file name.

- **Save all pages from the same image in one file** – all the pages originating from a single multi-page image file are saved to a single file.

- **Create a new file at each blank page** – each page group is saved to a separate file. These groups are separated by blank batch pages.

- **Create a single file for all pages** – all (or all selected) batch pages are saved to a single file.

# Saving the Recognized Text in RTF and DOC Formats

Layout retention modes are set on the **Formatting** tab in the **Options** dialog (menu **Tools>Options**).

---

**Note:** When you save the text in RTF or DOC formats, the fonts used are those set on the **Formatting** tab in the **Options** dialog (menu **Tools>Options**) or those you set during text editing in the **Text** window.

---

**Tip:** If you prefer editing the recognized text in Microsoft Word rather than in the FineReader **Text** window, you may still have uncertain characters highlighted. To do this, select the **With background color** and/or the **With text color** items on the **RTF/DOC** tab in the **Highlight uncertain characters** group. The saved file will have all the uncertain characters highlighted with the color you've set.

---

# Saving the Recognized Text in PDF Format

## Document layout retention options:

1. **Text and pictures only** – only the recognized text and the pictures in it are saved.

2. **Text over the page image** – the entire image is saved as a picture. Text areas are saved as text over the picture.

3. **Text under the page image** – the entire image is saved as a picture. The recognized text is put under it. This option is useful if you export your text to document archives: the full page layout is retained and the full-text search is available if you save in this mode.

**To set these options:**

1. Select the **Formats Settings** item in the **Tools** menu. The **Formats Settings** dialog will open.

2. Set the options you need on the **PDF** tab in the dialog.

---

**N o t e :** A special **Replace uncertain words with images** option is available if you use the **Text and pictures only** or the **Text over the page image** mode. If you set this option, all uncertain words will be replaced with their images. You may set this option on the **PDF** tab in the **Formats Settings** dialog.

---

## Font use modes

You should set the font use mode if you save in PDF format. If you are deal-ing with texts written in a code page different from the Latin code page (e.g. Cyrillic, Greek, Czech, etc.), set a Type 1 font use mode and the Type 1 fonts themselves as well. Adobe Type Manager must have these fonts avail-able.

1. **Use Acrobat Reader standard fonts** – PDF file refers to the standard system fonts Times, Helvetica and CourierNew.

2. **Refer to Type 1 fonts** – only references to Type 1 fonts are registered in the PDF file, the fonts themselves aren't embedded.
   Fonts referred to by the PDF file must be installed and available with Adobe Type Manager.

3. **Embed Type 1 fonts** – the Type 1 fonts are embedded into the PDF file.

---

**N o t e :**

1. If you don't have Type 1 fonts, you may use **Use Acrobat Reader stan-dard fonts** mode to save all documents. But the characters from non-Latin code pages, such as Cyrillic, Greek, Czech etc. will be properly displayed only in Acrobat Reader 3.x on Windows platforms. The characters from Latin code page are will be correctly displayed in any Acrobat Reader version.

2. In the **Refer to Type 1 fonts** mode only references to the fonts in use are registered in the PDF file (the fonts themselves are not actually embedded in the PDF file as they are in the **Embed Type 1 fonts** mode), so the PDF file occupies less disk space than the same file saved in the **Embed Type 1** fonts mode.

3. Saving in the **Embed Type 1 fonts** mode enables other users to view, edit and print the document set in the original fonts even if these fonts are not installed on their computers.

---

### To set the Type 1 fonts to be used:

- Click the **Type 1 fonts** button on the **PDF** tab in the **Formats settings** dialog (menu **Tools> Formats Settings**).

**Note:** You must set all the font types: serif, sans serif and monospaced in the **Type 1 fonts** dialog.

# Saving the Recognized Text in HTML Format

Layout retention modes are set on the **Formatting** tab in the **Options** dialog (menu **Tools>Options**).

**Note:** When you save the text in HTML format, the fonts used are those set on the **Formatting** tab in the **Options** dialog (menu **Tools>Options**) or those you set during text editing in the **Text** window.

**To retain pictures in an HTML file:**

- Set the **Keep pictures** option on the **Formatting** tab in the **Options** dialog (menu **Tools>Options**)

**Note:** Pictures are saved into separate \*.jpg files.

## HTML formats available

1. **Full (uses CSS, requires Internet Explorer 4.0 or later)** – if you set this option, the new HTML format – HTML 4 – is used. It supports any type of document layout retention (the actual retention type depends on the options set on the **Formatting** tab in the **Retain layout** group); the built-in style sheet is used.

2. **Simple (compatible with all (Internet) browsers)** – if you set this option, the HTML 3 format is used. The document layout is retained approximately: the first line indent is not retained and the font size is saved approximately (the HTML 3 format supports only a limited number of font sizes, so FineReader will choose which HTML 3 format font size corrsponds to the actual font size of your text). This HTML format is supported by all browsers (Netscape Navigator, Internet Explorer 3.0 and later).

3. **Auto (save Full and Simple formats in a single file with autoselection depending on browser type)** – both formats (**Simple** and **Full**) are saved to the same file. The browser you use will select the format it supports.

**To set the HTML format you want:**

- Click the radio button you need on the **HTML** tab in the **Formats Settings** dialog (menu **Tools>Formats**) in the **Formats** group.

---

**N o t e :**  The application detects the code page automatically. To change the code page select the one you need in the **Code page** field on the **HTML** tab in the **Formats Settings** dialog.

---

# Saving the Page Image

1. Select a batch page.

2. Select the **Save Image As** item in the **File** menu. The **Save as** dialog will open.

3. Select the disk and the folder you want to save the file to, and select also the file format.

---

**N o t e :**  If you want, you may save only some image areas enclosed by blocks. To do it, select the **Save only selected blocks** item in the **Save Image as** dialog.

---

4. Specify the file name.

5. Click **OK.**

---

**N o t e :**  You may save several images to a single file (a multi-page TIFF). To do it:
1. Select the images you want in the **Batch** window.
2. Select the **Save Image As** item in the **File** menu. Select the TIFF format and the **Save as multipage image file** option.

---

C h a p t e r   9

# Working with Batches

The batch is the main ABBYY FineReader data depository: the scanned images, the recognized text and other data are all kept in the batch. The majority of FineReader options are set for a batch: scanning, recognition, saving options, etc. User patterns, user languages and user language groups are batch "property" too. When you create a new batch, you may use either the default batch settings, the settings of the current batch or the settings saved in an *.fbt file.

## Chapter contents:

- General information on working with batches
- Creating a new batch
- Opening a batch
- Adding images to a batch
- Batch page number
- Closing a batch
- Deleting a batch

# General Information on Working with Batches

When FineReader starts for the first time, it opens the batch located in the FineReader folder. You may work with this batch or create a new one. A batch may contain up to 9999 pages.

---

**Tip:** You may find it useful to save the "same" pages in the same batch. That is, if you keep, say, all pages from the same a book in the same batch, or keep texts in the same language or images with the same layout in the same batch, and so on, you will be able to find your work much easier.

---

The **Batch** window displays the list of pages of the open batch. To view a page, just click its icon or number. All files related to this batch page will open in their respective windows, i.e. text file (if the page has been recognized) in the **Text** window and image file in the **Image** window.

**There are two main ways of displaying pages in the Batch window:**

| Batch View | Description |
|---|---|
| **Thumbnails** | In this view, the batch pages are displayed as thumbnails. A thumbnail is just a greatly zoomed-out image of the original page. Additional icons appear on the thumbnails as you process the image, telling you that a particular action has been performed on it, i.e. recognition, saving etc. You may find this viewing facility good for opening a particular batch page: you see a small page image in the batch window so you will easily find the page you need. To open an image just click its thumbnail. |
| **Details** | In this view, the batch window displays detailed information about each batch page; you may sort the page list by any feature you need. The **Details** view is convenient if you are dealing with a large batch, as the batch window displays a much greater number of pages in this view than in the **Thumbnail** view. Double-click a page to open it. |

## To choose the page view in the Batch window:

● Right-click the **Batch** window and select the **View>...** item in the local menu.

You may also tune each **Batch** window view, i.e. choose the displayed features, the way the pages are sorted etc. To do it:

● Right-click the **Batch** window and select the **Batch View>Customize**

item in the local menu. A dialog will open. Set the options you like on the **Thumbnails** and **Details** tabs of the dialog.

You may select several pages or a number of pages in a row:

- **To select a number of pages in a row**, hold down **SHIFT** and click the first and then the last page of the group you want to select.

- **To select several pages**, hold down the **CTRL** key and click the pages you need.

- **To select all the batch pages**, activate the **Batch** window and choose **Select All** item in the **Edit** menu or press **CTRL+A**.

# Creating a New Batch

### To create a new batch:

1. Select the **New Batch** item in the **File** menu. The **Create New Batch** dialog opens.

2. Select or create a folder for the new batch in the **Create New Batch** dialog.

3. Select the **Batch Template** field and choose one of the following values depending on the settings you want to be applied to the new batch: **Default settings** – to apply default settings, **Current batch** – to apply the current batch settings, **Batch template** (.fbt) – to apply settings earlier saved to a special file.

---

**N o t e :** To save the batch settings in a file click the **Save** button on the **General** tab (menu **Tools>Options**). A **Save Batch Template as** dialog will open. Specify the file name in it. The following options are saved in this file: the options set on **Recognition, Scan/Open Image, Formatting** and **Check Spelling** tabs, the options set on other **Formats Settings** dialog tabs. User languages, user language groups and user patterns are also saved in this file. If you'd like to switch back to the default settings just click the **Use defaults** button on the **General** tab. To load settings click the **Load** button on the **General** tab and select a FineReader batch template (*.fbt) file containing the options you need.

---

# Opening a Batch

1. Select the **Open Batch** item in the **File** menu. The **Open Batch** dialog will open.

2. Select the folder containing the batch you need in the **Open Batch** dialog. When you open a batch, the batch you've been working with previously is automatically closed and saved.

FineReader opens the last batch you worked with automatically at startup.

# Adding Images to a Batch

1. Select the **Open Image** item in the **File** menu or press **CTRL+O**.

2. Select the image(s) you want to open in the **Open Image** dialog. FineReader will add the image to the open batch and copy the image to the batch folder.

# Batch Page Number

All batch pages are numbered. One batch may contain up to 9999 pages. The page number is displayed in the batch.

### To renumber pages:

1. Select a single page or several pages.

2. Select the **Renumber Pages** item in the **Batch** menu.

3. Set the new number for the first selected page (the page with the least number).

---

**Note:**
1. If you want to renumber all the batch pages, select the **All Pages** item in the **Renumber Pages** dialog.
2. If you want to renumber only a part of the batch:
   - Select the pages you want to renumber in the **Batch** window.
   - Select the **Selected pages** item in the **Renumber Pages** dialog.
3. If you want the selected pages to be renumbered continuously, select the **Continuous page numbering** option. For example: the page numbers were 2,5,6, the 1 was set as the first number; after renumbering the page numbers will be 1,2,3. Otherwise (i.e. in case the **Continuous page numbering** option has not been set) the page numbers will be 1,4,5.

---

# Closing a Batch

1.  Select the **Close Batch** item in the **File** menu.

The batch will be saved automatically when you close it.

# Deleting a Batch

---

**N o t e :** Deleting a batch means deleting all its contents, i.e. all its pages (images and text) together with all additional files you've created during your work with the batch: user patterns, user languages etc. That is, you delete all the contents of the batch folder.

---

1.  Select the page(s) you want to delete in the **Batch** window.
2.  Select the **Delete Page** item in the **Batch** menu or just press **DEL**.