Spelling Correction System Acknowledgments

International CorrectSpell™ Catalan spelling correction system © 1995 by INSO Corporation. All rights reserved. Adapted from Catalan word list © 1992 Universitat de Barcelona. Reproduction or disassembly of embodied algorithms or database prohibited.

International CorrectSpell™ Czech spelling correction system © 1995 by INSO Corporation. All rights reserved. Adapted from word list supplied by Jan Hajic. Reproduction or disassembly of embodied algorithms or database prohibited.

International CorrectSpell™ Danish spelling correction system © 1995 by INSO Corporation. All rights reserved. Portions adapted from The Orthographical Dictionary, 5th Ed. 1988, by the Danish Language Council. Reproduction or disassembly of embodied algorithms or database prohibited.

International CorrectSpell™ Dutch spelling correction system © 1995 by INSO Corporation. All rights reserved. Reproduction or disassembly of embodied algorithms or database prohibited.

International CorrectSpell™ English spelling correction system © 1995 by INSO Corporation. All rights reserved. Reproduction or disassembly of embodied algorithms or database prohibited.

International CorrectSpell™ Finnish spelling correction system © 1995 by INSO Corporation. All rights reserved. Adapted from word list supplied by the University of Helsinki Institute for Finnish Language and Dr. Kolbjorn Heggstad. Reproduction or disassembly of embodied algorithms or database prohibited.

International CorrectSpell™ French spelling correction system © 1995 by INSO Corporation. All rights reserved. Adapted from word list supplied by Librairie Larousse. Reproduction or disassembly of embodied algorithms or database prohibited.

International CorrectSpell™ German spelling correction system © 1995 by INSO Corporation. All rights reserved. Adapted from word list supplied by Langenscheidt K.G. Reproduction or disassembly of embodied algorithms or database prohibited. © Licensee and others. 1995.

International CorrectSpell™ Greek spelling correction system © 1995 by INSO Corporation. All rights reserved. Reproduction or disassembly of embodied algorithms or database prohibited.

International CorrectSpell™ Hungarian spelling correction system © 1995 by INSO Corporation. All rights reserved. Portions of technology and word list supplied by Morphologic. Reproduction or disassembly of embodied algorithms or database prohibited.

International CorrectSpell™ Italian spelling correction system © 1995 by INSO Corporation. All rights reserved. Adapted from word list supplied by Zanichelli S.p.A. Reproduction or disassembly of embodied algorithms or database prohibited.

International CorrectSpell™ Norwegian spelling correction system © 1995 by INSO Corporation. All rights reserved. Reproduction or disassembly of embodied algorithms or database prohibited.

International CorrectSpell™ Polish spelling correction © 1995 by INSO Corporation. All rights reserved. Portions of technology and word list supplied by Morphologic. Reproduction or disassembly of embodied algorithms or database prohibited.

International CorrectSpell™ Portuguese spelling correction system © 1995 by INSO Corporation. All rights reserved. Portions adapted from the Dicionario Academico da Lingua Portuguesa. © 1992 by Porto Editora. Reproduction or disassembly of embodied algorithms or database prohibited.

International CorrectSpell™ Russian spelling correction system © 1995 by INSO Corporation. All rights reserved. Reproduction or disassembly of embodied algorithms or database prohibited.

International CorrectSpell™ Spanish spelling correction system © 1995 by INSO Corporation. All rights reserved. Adapted from word list supplied by Librairie Larousse. Reproduction or disassembly of embodied algorithms or database prohibited.

International CorrectSpell™ Swedish spelling correction system © 1995 by INSO Corporation. All rights reserved. Reproduction or disassembly of embodied algorithms or database prohibited.

Table of Contents

Welcome

Chapter 4 Working with Documents 45

Chapter 5 Improving Recognition Accuracy 69

Welcome

Welcome to Recognita Plus 5.0, a multi-lingual Optical Character Recognition (OCR) program running under Windows 95, Windows 98, Windows NT 4.0 and Windows 2000. The program enables you to convert your paper documents or image files to computer editable text in an easy and convenient way. The following documentation has been provided to help you learn about Recognita Plus.

This Guide

This guide is intended to give you a basic knowledge of Recognita Plus. It includes installation and setup instructions, gives you a general idea on optical character recognition and of what this software can do for you. It shows you the typical steps for processing your documents. The guide does not, however, cover all the particulars or possible functions.

Electronic Online Help

Going into more detail, the Electronic Online Help provides exact documentation of all the features, settings and procedures and gives answers to the widest possible range of questions.

Tips of the Day

Each time you start the program the Tip of the Day window pops up (unless you disable it), displaying useful hints about different features of Recognita Plus. By reading these ideas, you will be able to exploit more and more of Recognita Plus's capabilities.

Supported Scanners

See the section "System Requirements" in Chapter 1 for information on the scanner(s) you are going to use with Recognita Plus.

Installation and Setup

In this chapter you will find information on the following topics:

- System Requirements
- Installation
- Setting up your Scanner for Recognita Plus
- Registration

System Requirements

You need a configuration with at least the following characteristics to install and run Recognita Plus:

- IBM compatible PC with Intel Pentium or equivalent processor.
- Microsoft Windows 95, Windows 98, Windows NT 4.0 or Windows 2000 operating systems.
- 8 MB of memory (RAM) for Windows 95 and Windows 98 (16 MB recommended),

  16 MB of memory (RAM) for Windows NT 4.0 and Windows 2000 (32 MB recommended).
- 35 to 45 MB free space on your hard disk, depending on the installation options you choose. To store your work with Recognita Plus, you need a lot more space, especially when creating long multi-page documents and having images embedded in your Recognita Documents.
- If you want to scan your paper documents, you need a supported scanner with 300 or 400 dpi resolution. For information on directly supported scanners, refer to the files scan_xxx.rtf supplied with the program (xxx is a language dependent part of the file name, it is *eng* for English, *ger* for German, etc). You can access the file contents in your setup language through the shortcut "Recognita Scanner Drivers" in the Recognita Program Group. More scanner information is provided on our Web-site *www.caere.com/recognita*. For information on scanners accessed through Caere's Scan Manager, use the shortcut to the "Scan Manager Setup Notes" in the Recognita Program Group. You can use Recognita Plus without a scanner to process image files.
- VGA monitor (preferably with more than 256 color support for handling color images).
- Mouse or other pointing device.
- CD-ROM drive at installation time.

Installation

You are guided through the installation with clear instructions at each step. First, please exit any applications that may be running or were auto-started.

Important: Under Windows NT 4.0 and Windows 2000, you need administrator privileges to perform installation.

To install Recognita Plus:

1. Insert Recognita Plus 5.0 CD-ROM in your CD-ROM drive. Wait for *setupocr.exe* to start automatically. If it does not start, locate your CD-ROM drive either in the Windows Explorer or in the Browse dialog box of the Start Menu's Run command and run *setupocr.exe* from your CD root.

2. First, you are prompted to enter the CD-Key. You can find it on the back of the CD-ROM holder.

3. Recognita Setup Wizard takes over. Select an installation language and follow the instructions on screen.

4. Click on Next at each step of the installation if you have specified the settings that you were asked or click on *Back* to change any of the settings specified at an earlier step.

5. Click on *Finish* to complete the installation and have the necessary files copied to the folder you specified.

6. After these steps, you have control over the following settings, presented in a tabbed dialog box:
   - Program languages (i.e. the language used in menus, messages, etc.)
   - Help file languages
   - Text output converters
   - Dictionaries, used for spelling and Language Analysis
   - Direct connection to applications and integration into mailing systems.

Note: During installation Recognita Plus's Maintenance Setup program will also be added to the Recognita Plus 5.0 program group. You can use it later to make changes to the current Recognita Plus setup, for example add a new scanner driver or output text converter, enable a direct connection, etc. An uninstall facility will also be placed in the group.

Setting up your Scanner for Recognita Plus

Recognita Plus can access scanners in different ways. Using Caere's Scan Manager is the preferred method, it is set as default during installation.

Scan Manager is a regularly updated software package from Caere Corporation providing consistent access to a wide and increasing number of scanners. Scan Manager is automatically installed as the last step of Recognita Plus setup. It displays a dialog box, offering a list of scanner brands. Use this only if you want to scan through Scan Manager or set 'No scanner'. The first item in its list is (Generic). Choose this to set "No scanner" or a generic TWAIN or ISIS interface. In these last two cases you should check whether the delivery settings are suitable. To choose a named scanner, click the brand to get a list of models. Select the one(s) desired. Scan Manager usually accesses the chosen scanner through TWAIN, but makes all the necessary settings automatically.

Scan Manager's setup program adds an icon to the Windows Control Panel. You should click that icon to change the installed scanner or any of its settings.

If you do not have a scanner, you can still use Recognita Plus to process image files scanned by other scanning software or arriving by fax boards and through E-mail. In this case you must remove Scan Manager from its default position, or select (Generic)/No scanner.

If you experience scanner difficulties see the next topic "Changing the Scanner Setup"

Changing the Scanner Setup

Before changing the scanner setup make sure your scanner runs with the software provided by the scanner manufacturer. During setup please have your scanner turned on. You can modify your scanner settings by using Recognita Maintenance Setup.

You can access a scanner by choosing:

- A specific scanner offered by Caere's Scan Manager program.
- A generic scanner driver offered by Caere's Scan Manager program.
- A scanner driver supplied with Recognita Plus.
- One of the TWAIN drivers supplied with Recognita Plus.

The following scanner setup dialog box is displayed by Recognita Maintenance Setup:
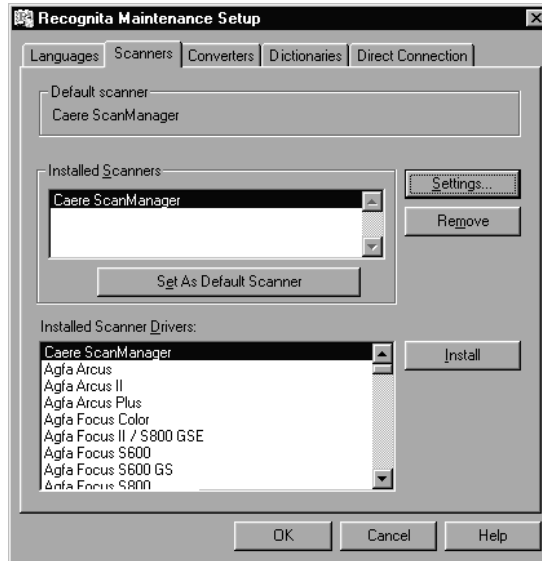
Scan Manager appears at the top of the list of Installed Scanner Drivers. It is automatically placed in the Installed Scanners panel and set as Default. Keep it there if you wish to use Scan Manager, and specify a scanner in its dialog box when it appears. If you do not want to use Scan Manager, either remove it or add one or more Recognita-supplied direct drivers, setting one as the default. The following topics explain how to setup Recognita scanner drivers in case of problems with Scan Manager.

To setup your scanner:

1. Turn on your scanner.

2. Select the scanner model from the list of the installed scanner drivers.

3. Click on *Install*. The driver name appears in the list of the installed scanners. If you have more scanners connected to your computer, you can install drivers for all of them in the same way.

4. A dialog box with the factory default settings of the scanner is displayed.
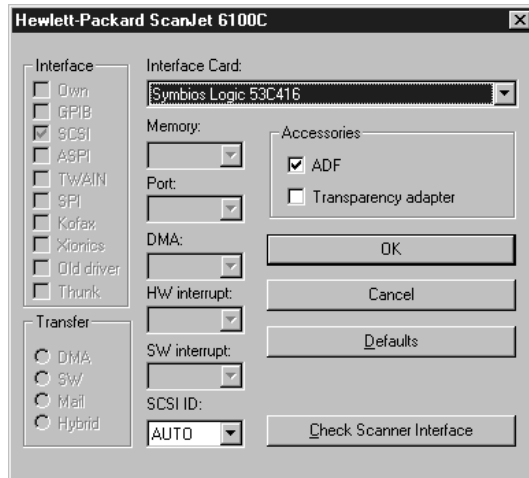
It shows settings such as Port addresses, Interrupt values, Interface cards, etc. Any grayed items are not needed for the current scanner. Check the values are correct. Specify whether an automatic document feeder (ADF) or transparency adapter is attached to the scanner.

5. Click on *Check Scanner Interface* to run a check on your configuration, to see whether all the information supplied is correct. If not, a message will advise which item needs attention. This might be a needed interface card not detected, an incorrect port address, etc. If you cannot immediately solve the problem, continue with setup, then consult the file *scan_xxx.rtf* to get a list of all factory default settings and run Maintenance Setup to change the scanner values as necessary.

6. Click on OK to return to the main scanner panel.

7. If you have installed more than one scanner, select one to be used currently, and click on *Set As Default Scanner*. You can change the default scanner later by running Recognita Maintenance Setup.

8. To remove an installed scanner, select it from the list of the installed scanners and click on *Remove*.

Setting up TWAIN Compliant Scanners

TWAIN is a standard interface for image capturing devices. Most scanner manufacturers provide TWAIN compliant drivers for their scanners.

Recognita Plus has its own scanner specific drivers for many scanner models and supports TWAIN.

If you have a TWAIN compliant scanner installed on your computer, you can choose from TWAIN specific entries during installation and when running Recognita Maintenance Setup. The scanner driver list contains one generic entry for TWAIN:

TWAIN: Basic Driver

and one or two items for each installed TWAIN compliant scanner in the form:

TWAIN: <data source name>

Two items are displayed if both 16 and 32 bit data sources have been installed under Windows 95 or 98. Always select a 32 bit driver if available.

Note: If your TWAIN compliant scanner model's name also appears on the driver list as a separate item (without the prefix "TWAIN"), you may choose it (and we recommend this) to be used by Recognita Plus through its own scanner driver.

Choosing a "TWAIN:<data source name>" entry:

The <data source name> contains the product name of the given data source (it is very often different from the scanner's actual model name). We suggest you choose this rather than the TWAIN: Basic Driver. If this is chosen, scanner settings can be set in Recognita Plus's user interface.

Choosing the "TWAIN: Basic Driver" entry:

The TWAIN: Basic Driver should be used only if you have problems with scanning when using a TWAIN:<data source name> type driver. If the TWAIN: Basic Driver is chosen, scanner settings can be set on the data source's own user interface, which appears when scanning is started from Recognita Plus.

When you complete *step* 5. or *step* 6. of setting up your scanner (see page 9), a Select Source dialog box appears with the list of the installed data sources' names. These names are identical to the ones in the "TWAIN:<data source name> type entries, but this time they appear without the prefix "TWAIN". You should specify here which one you want to use through the Twain Basic driver of Recognita Plus.

Other TWAIN issues:

The user interface of a TWAIN data source might offer settings unsuitable for OCR purposes. These can be for example extreme resolution values, halftone (dithered) image output and so on. Please avoid these for best results.

In some cases Recognita Plus might fall back to using TWAIN: Basic Driver despite your selection of a TWAIN: <data source name> driver. This is not an error, and happens if Recognita Plus detects that it cannot control all necessary scanner settings. Remember that in this case the scanner parameters can be set on the data source's own user interface. If you use the TWAIN: Basic Driver, you may try to enable the automatic document feeder handling mechanism of the data source. To do this, enter the line:

```
AdfHandling=1
```

into the SCANNER.INI file in the Recognita Plus folder. If this works, the data source's user interface will be displayed before the first page of a stack only. Otherwise it appears before each page.

Special Scanner Issues under Windows 95 and 98

If you get an error message during scanning **under Windows 95 or 98** and it is not likely that it is a real scanner error, you should insert the following line into your CONFIG.SYS file right after the HIMEM.SYS and EMM386.EXE entries:

```
DEVICE=<Recognita path>\RSDBUF.EXE [/8]
```

where <Recognita path> is the full pathname of Recognita Plus. Note that you must not use the DEVICEHIGH command in this line. This driver allocates buffers in the conventional memory for Recognita scanner drivers. If the /8 switch is given, less memory is allocated (8k).

Do not use the /8 switch for the following scanner types:

- Ricoh RS632 with ISI-8 interface card
- Siemens scanners
- Lightscan 400P
- Pentax DS6, DS10
- Mitsubishi MH216CG
- AVision scanners
- Dextra Reader
- Genius FastReader
- Mouse Systems PB/Reader
- Targa TS 30n, TS 600C, TS 800C

Registration

Registered customers of Recognita Plus 5.0 will:

- have access to our technical support services
- receive the latest information about new and improved Recognita products
- get upgrade offers at special prices.

Unregistered users are prompted to register periodically when Recognita Plus is started. Once you register you will not be prompted any more.

As a result of registration, you will get your *registration number*, which must be entered in the appropriate textbox of the Recognita Registration Wizard.

To get your registration number:

1. Choose *Registration…* from the Help menu to start the Recognita Registration Wizard. This program is also started the first time you start Recognita Plus.

2. Click *Next* on the introductory window. The following window appears:

3. To register, choose one of the three methods offered. Click on each to see how each method works.

Online

If you select *Online* and click *Next,* you are guided to our Registration Web Page where you can fill in an electronic form and receive your registration number immediately. Then switch back to the Registration Wizard, enter the number and click *Next* to verify it.

Offline

If you select *Offline* and click *Next,* the Registration Wizard will provide an electronic form. Fill it in, clicking *Next* for each new page. When you click *Register,* the program will first search an e-mail connection, then a fax modem. It will inform you which sending method was used. If neither was successful, it will print the form for you. If a printer is not available, it will invite you to save the form to disk. Please fax or post the form or use the registration card enclosed. Your registration number will be sent to you by e-mail, fax or post. Use *OK* to exit the Registration Wizard.

Phone

Phone registration is also available in some countries (currently the Czech Republic, Germany, Hungary, Poland and Sweden). Click on *Phone* and use the drop-down listing to see the number to use. Please be ready to dictate your serial number. If possible, phone with the Registration Wizard screen still active so you can enter your registration number and press *Next* to test it immediately.

To complete registration:

1. Choose *Registration…* from the Help menu to start the Recognita Registration Wizard if it is not running.
2. Move to the Registration method panel and enter your registration number in the textbox provided.
3. Press *Next* to have the number verified and to complete the registration process.
4. Note the number in a safe place; we recommend the space provided at the end of this Users' Guide.

Introduction to Recognita Plus

Have you ever been key-bored? Well, if the answer is no, then you are among the lucky ones, and it is not likely you'll ever be. However, if the answer is yes, then you probably know how tiresome retyping your printed documents can be. But why waste your precious time if a solution to this problem is near at hand.

Recognita Plus – as you might already have guessed – is the solution. This omnifont and multi-lingual OCR software converts your paper documents with the greatest ease and accuracy into computer editable form. As soon as you begin to use it, you will be convinced that this software really means the end of an era – the era of manual retyping.

In this chapter you will find information on the following topics:

- What is OCR All About?
- Processing Stages in Recognita Plus
- The Recognita Document
- Application and Document Windows
- The Electronic Online Help
- What's New Compared to Version 4.0
- Product Support

What is OCR All About?

Optical Character Recognition is the art or science of scanning printed documents and making their text content computer editable. The program examines the incoming shapes and decides which character each represents. Recognita Plus's technique is mainly based on contour analysis in which each character is defined by certain typical measurements or ratios of its contour elements. This has the advantage of making recognition *omnifont*: much more independent of character size and font variations. As a supplement to its base algorithm, the program also uses Self Assertion Technology (SAT) which uses improved pattern matching. In addition, the OCR engine consults the Language Analysis module of the recognition language on the words being built from the recognized characters. These techniques together ensure optimum recognition.

A new level of accuracy is introduced with Recognita Plus 5.0. It is available for eleven major European/American languages. The program is equipped with two recognition engines, both with their own Language Analyst support. The two engines read texts in parallel and compare results. Where differences arise, certainty data from both engines are used to accept the best solutions. Tests on degraded documents have shown the number of errors can be reduced by up to 30%.

Processing Stages in Recognita Plus

People are not the same. Neither are the tasks they have to solve day-by-day. What they all may need to make their lives easier is a flexible tool, which can be tailored to their needs.

Recognita Plus is a versatile product which can be used in many ways to process single-page or multi-page documents. From step-by-step manual interaction to fully automated document processing, everything is possible. This guide does not cover all the possibilities but describes the most typical processing steps. Besides this, it draws your attention to settings that have an effect on how a document can pass through Recognita Plus. To learn about these settings please read the relevant topics in the online help of Recognita Plus.

Processing steps in Recognita Plus

1. Obtaining the source

This involves getting some input. It can mean *scanning* to create a digitized image of the document. It can mean *opening* an existing image file, either from Recognita Plus, the Desktop or Explorer, or taking the image attachment from an e-mail.

Scanning and image import can be in black-and-white, grayscale or color. Images are displayed as imported.

2. Image pre-processing

The program automatically prepares the acquired image(s) for optimum recognition by detecting and removing any skew and making sure orientation is correct. (You can pre-define orientation or leave the program to detect it.).

3. Decomposition, zoning

This involves finding information on the page and zoning it. The program automatically distinguishes graphics from text; text is classed as flowed text or a table. The program also decides a reading order for the zones.

Manual zoning is also possible. You can draw zones, modify their size, position, order and assign a recognition engine. Zone templates can also be applied.

4. Recognition

This is the heart of the operation. Here one or more of the six recognition engines is used, depending on the zone properties. The engines available are: ***Omnifont*** (most often used), Dot matrix, Checkmark, Barcode, Braille and Handprint (for numbers only). As a result of the recognition process, you get a Recognita Document with formatted, editable text. Typically, the processing stops after this step. If stopped, any page can be re-recognized with changed settings.

5. **Proofing, training, editing**

    These functions make up the correction phase and are controlled fully by the user. *Proofing* helps you find any problem areas, such as non-dictionary words or suspect characters. *Training* can be used to teach the program repeatedly misread characters. The built-in editor offers most normal word processor *editing* functions for correcting and formatting the text.

6. **Saving and exporting**

    You can save the text in a wide range of text formats with a formatting level of your choice. *Images* can also be saved in many popular image file formats. In addition to these, you can also save your work as a *Recognita Document,* containing both text and images, ready for further processing. Copying to the Clipboard and sending mail attachments are also possible. This step can be either manual or automatic.

The Recognita Document

A Recognita Document (file extension RCD) consists of pages which contain or are linked to the acquired *images* of your document and – if recognized – also contain editable *text*. *Data* related to the images and texts on the page are also stored. This file format is unique to Recognita Plus. Each character or recognized element is linked to the corresponding part in the original image so that proofing, verifiers and the training module can function.

Recognita Document files can be saved and later reopened by the program, providing the basis for deferred processing, that is, for example, doing scanning one day, recognition the next, full-facility proofing and text saving on the third or any later day. You should retain your Recognita Document files as long as you expect that you might want to save some or all of their contents (either text or images) in an output format Recognita Plus supports.

Application and Document Windows

Recognita Plus can handle more than one document at a time. Recognita Documents are displayed in document windows in the working area of the main application window. A typical document window in its maximized state is shown below.

Browser List

Text pane (Built-in editor): for proofing, editing and training

Main toolbar

Options toolbar

Editing toolbar

Browser images

Pages Browser pane: for navigating easily in a multi-page document

Click here to display the image Overview window

Image pane: for displaying the original image and doing zoning

To get information on the various screen elements, their purpose and usage, consult the context sensitive help.

The Electronic Online Help

Recognita Plus has a comprehensive help system: both *context sensitive* and *general*. You can use it to get detailed information on features, settings and procedures.

The Help Menu
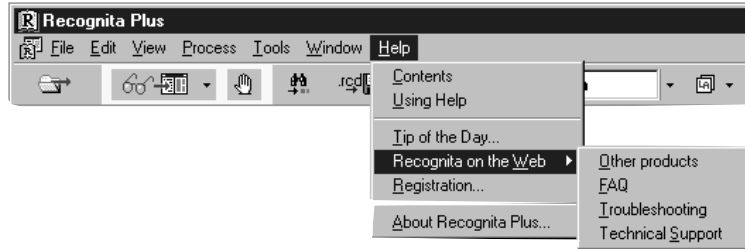
- Choose *Using Help* to get an overview of how to use help.
- Choose *Contents* to display information organized by category, to select an item from the help index, or to search for specific words and phrases in help topics rather than searching by category.
- Choose any of the menu items from the submenu *Recognita on the Web* to navigate to a Web page of Recognita Corp. and get the latest information on products, troubleshooting, supported scanners etc.
- Choose *Tip of the Day* to get useful ideas and suggestions for using Recognita Plus. The Tip of the Day window is displayed each time Recognita Plus is started unless you disable it.

The Context Sensitive Help system:

- ToolTips give short explanations on a screen element, typically a toolbar button. They appear if the cursor stays still over an item for a second or so.
- Status bar messages give explanations of a menu item or toolbar button. They appear if a menu item is highlighted or a button is being pressed.
- Put the cursor on a menu item and press F1. It works for all menus but this is the only way to get help on a context menu item, i.e. an item in a menu appearing when the right mouse button is clicked.

- Click on the Help button then on any menu item, tool or area to get help on its purpose.
- Dialog boxes have their own help tool, top right. Click on it, then on any part of the dialog box to get help on its purpose.
- Some dialog boxes have a Help button besides the small question mark tool. Click on it to have an overview on the purpose of the dialog box.

  In this User's Guide, function key or key combination symbols in the left margin inform you if a command mentioned in the text is also accessible by the key(s) shown.

  In the Reference section of the online help, you can find keyboard guides, a summary of cursor shapes, settings and language lists and a glossary.

What's New Compared to Version 4.0

Recognition

- Maximum accuracy from dual-engine recognition available for 11 languages. Two OCR engines read the text in parallel, both using dictionary support. They compare solutions and confidence levels for real accuracy gains, especially on degraded documents.
- Choose from 6 recognition levels, from fastest to most accurate: with one-, two- or three-step reading, with or without support of a Language Analyst and single or dual-engine recognition.
- A new OCR-specific deskew algorithm yields greater accuracy.

Image handling

- Color and gray images can be scanned, displayed, printed and exported. Graphics zones in recognized text files can also contain color images. Mixed image types (black-and-white, gray, color) can now be saved to a single multi-page image file.
- A preview feature makes it easier to navigate and find required image files.

- The program includes Caere's Scan Manager 5.0, opening the way to much wider scanner support.

Languages

- Cyrillic alphabet support is introduced, with ten languages offered: Bulgarian, Byelorussian, Chechen, Kabardian, Macedonian, Moldavian, Ossetian, Russian (with dictionary support), Serbian and Ukrainian.
- The language list can be customized. On delivery, the languages with dictionary support appear. A second list can be invoked, presenting all 114 supported languages. Languages can be added, removed or reordered as desired.

Proofing and editing

- The static pop-up verifier can be replaced by a dynamic one which remains open and tracks the editing position, with the current character always centred in the pop-up window.
- The side-by-side verifier is now referred to as the Image pane verifier.
- The Find facility can be set to find whole-word occurrences only.
- The decimal separator in tables has become user-definable.

Processing

- The Stop for (Re)zoning feature can be turned off or on while processing is paused, allowing the zoning method to be changed midway in a document.
- A two-page template facility makes it easier to process two-page forms or books.
- The Revert to Saved facility remains available for selected pages in the Browser's context menu, but also appears in the Main menu, where it functions for the whole document.
- Improved saving support for exporting recognized texts to MS Word 97.

Improved support for the visually handicapped

Braille recognition

Braille can be set as the General zone type for a whole document. Auto-decomposition places single whole-page recognition zones. Manual zone drawing is possible, but all zones in the document must be for Braille recognition; numbers-only zones are permitted. Output will be the editable text equivalent of the Braille text.

General modifications

The Text, Image and Browser panes can be maximized by Ctrl+1, Ctrl+2, Ctrl+3 respectively. Ctrl+4 restores the current pane. The focus can be moved from one pane to the next by F6 and toggled between the Browser's list and pages by Tab. When a single page is selected in the Browser list, F2 allows text entry into the Note field. Training suggestions can receive the focus, making them available to screen readers. Direct connections can be activated by Hot Keys.

Specific modifications

The following modifications can be invoked by starting Recognita Plus from a command line with the /blind option: In the View/Columns dialog box, checkmarks are replaced by YES/NO texts. Edit box displays use Windows Code Page characters, not Recognita Fixed Fonts. The six-position Speed/Accuracy slider in the Options/Accuracy panel is replaced by a drop-down listing. That means all these controls can be handled by a screen reader.

Product Support

Please register your copy of Recognita Plus to be eligible for product support. If you have any questions about Recognita Plus and you don't find the answer in this guide or in the online help, you can get help from the following services:

WWW home page

If you visit our home page, you can get information on other Recognita products, troubleshooting techniques, updates and answers to Frequently Asked Questions (FAQ). Access this:

- From the Help menu
- At www.caere.com/recognita

You can send your technical questions through the Internet on a form available on our homepage.

Telephone service

You can send a fax or call our technical support staff on the following numbers:

- Fax: (36 1) 452-3710
- Tel.:   (36 1) 452-3706

Our technical support staff is ready to give you the support you need to get the most from your Recognita Plus software. When you call, please have the following near at hand:

- Recognita Plus version and registration number
- The make and model of your scanner
- The names and version numbers of the other scanning software you use with your scanner
- The amount of memory (RAM) on your system
- The amount of free hard disk space on your Windows drive

- The amount of free hard disk space on the drive where your temporary files are stored. To list system settings including the path where temporary files are stored, type SET at the command prompt and press Enter. The TEMP keyword shows the path in question.

  Free hard disk space is indicated on the status bar of the Windows Explorer. Select the drive letter of the hard disk in question.

Should you experience an error using Recognita Plus, please:

⇨ Make an exact record of any error messages.

⇨ Record the steps to reproduce the error.

⇨ If possible, save your zone pattern to a template file which you could send to us together with the problem image(s).

When calling by phone, please have Recognita Plus running on your computer if possible.

Processing Documents

This chapter provides information on processing your documents with Recognita Plus. There are different ways to scan, recognize, correct and save a document. Depending on the quality and number of pages to be processed, the time you intend to spend on the work, the required accuracy and the preferred proofing method, you may choose from many different possibilities. You can control the processing stages step-by-step or choose fully automatic processing.

In this chapter you will find information on the following topics:

- Overview of Processing
- Creating Documents
- Interrupting and Continuing the Process
- Recognizing Images in a Document
- Working with Documents
- Saving Documents, Text and Images
- Starting Recognition from Other Applications
- Processing and Saving without Display

Overview of Processing

The following diagram tries to summarize the main processing steps available in Recognita Plus.
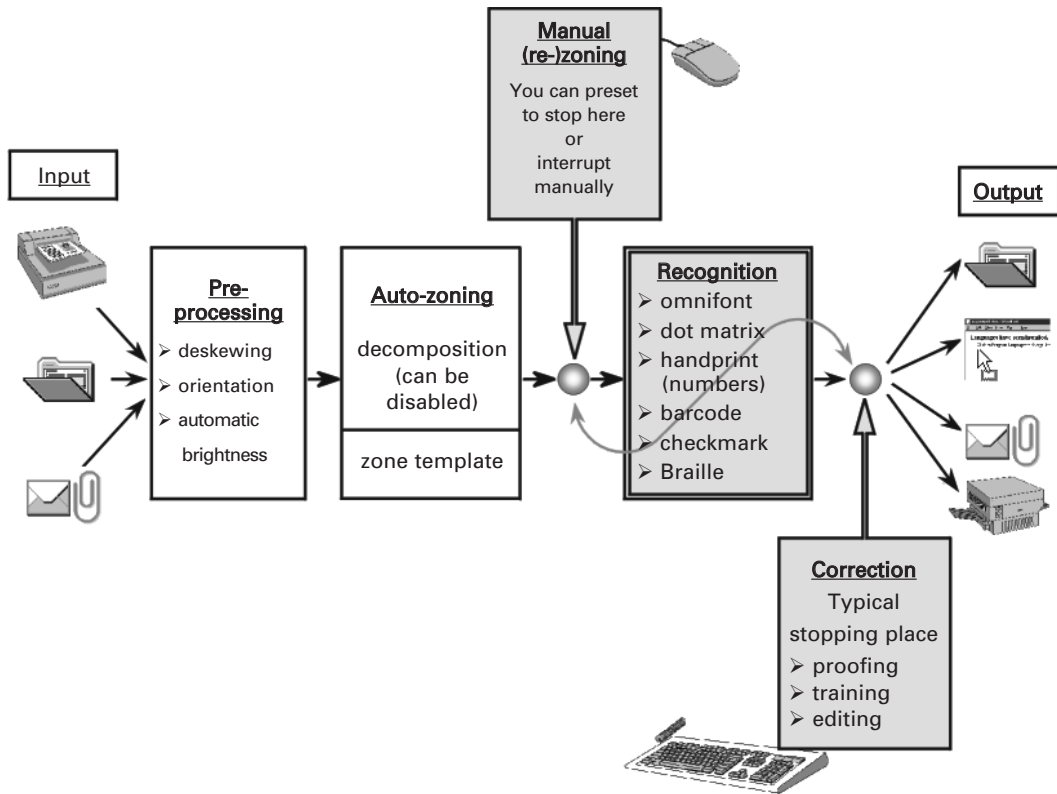
The process stops (or can be stopped) at the points indicated and the program allows different user interactions. Re-recognition of images with modified settings is possible. You can also save the current state of the document to a Recognita Document file at any time. Re-opening it later is the key to deferred processing.

In addition to the possibilities shown above, Recognita Plus offers the unique feature *Save Without Display*. If it is enabled, the program processes scanned pages or image files and saves the results (images, text or Recognita Documents) fully automatically to a series of output files without user interaction.

Creating Documents

You can use the sample files shipped with Recognita Plus for the procedures starting from image files. To scan printed documents please choose some and have them near your scanner.

Note that there is no such a thing as an empty Recognita Document. A new document is always created by scanning printed materials or loading image files, and – if required – recognizing their contents. In general, images can be embedded in a Recognita Document file; image files can also be linked by path and name.

There are two basic methods of scanning/loading images:

- Scanning/loading *and* recognizing. This method results in a document with images *and* text.
- Scanning/loading only. This method results in a document with images only. Recognition can be carried out later.

When a process is started and there is at least one document open, the Next Document dialog box appears:
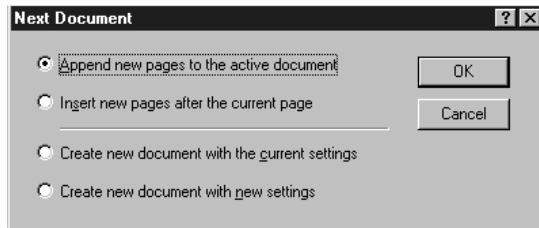
⇨ Choose one of the first two options if you want to add the new pages to the active document.

⇨ Choose one of the other two options if you want to create a new document. If you choose the last one, the Options dialog box will be offered to allow the new settings to be specified.

To load (and optionally recognize) image files:

1. If you want to use the toolbar to start processing, make sure the selected image source is *file* and not scanner. You can toggle between the two by clicking the leftmost button on the Main toolbar.

2. Start loading files with or without recognition.

With recognition:

⇨ Make sure the main processing button shows the image on the left and click on it or

⇨ Choose *Read>from File* in the Process menu.

Without recognition:

⇨ Make sure the main processing button shows the image on the left and click on it or

⇨ Choose *Scan>from File* in the Process menu.

The File(s) to Open dialog box appears.

You can add files to this list, e.g. from different paths
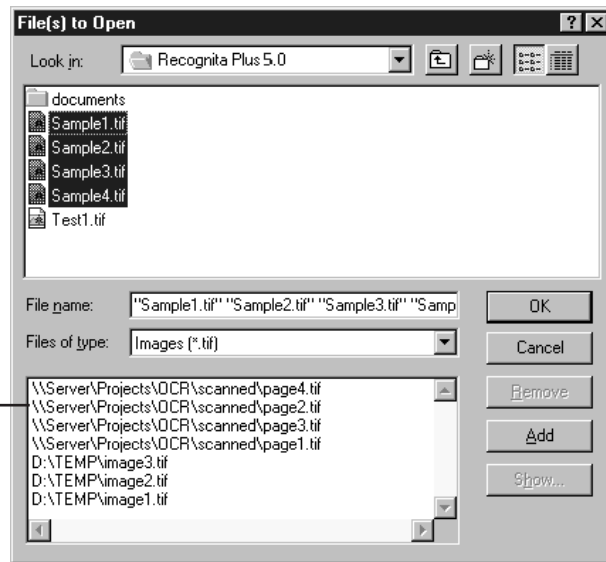
It will show the last used folder location. Select the files you want to recognize. Selected files plus files listed in the lower panel will be processed. Whenever a single file is selected in either panel, click *Show…* to see a quick preview image of the file.

Click *Add* to add the selected files to the list in the lower panel if you want:

• to process files from different folders

• to process the files in a specific order

3. Choose OK to start processing the files. Progress is indicated on the status bar. If recognition was selected, the progress of the OCR process is also indicated in an overview window showing the image.

4. At the end of the process, the first page processed will be shown.

To scan (and optionally recognize) paper documents:

1. If you want to use the toolbar to start processing, make sure the selected image source is *scanner* and not file. You can toggle between the two by clicking the leftmost button on the Main toolbar.

2. Place the page(s) to be scanned in your scanner. You can scan a stack of pages in one process if you have an automatic document feeder (ADF).

3. Start scanning with or without recognition.

   With recognition:

   ⇨ Make sure the main processing button shows the image on the left and click on it or

   ⇨ Choose *Read>from Scanner* in the Process menu.

   Without recognition:

   ⇨ Make sure the main processing button shows the image on the left and click on it or

   ⇨ Choose *Scan>from Scanner* in the Process menu.

4. Wait for the pages to be scanned in and processed. Progress is indicated on the status bar. If recognition was selected, the progress of the OCR process is also indicated in an overview window of the original image.

5. When no more pages are available, a dialog box appears, asking you if you want to scan more pages. Choose YES if you want to scan more pages into the same document or NO if you want to stop the process. Place the new page(s) into the scanner before choosing YES.

6. At the end of the process, the first page processed will be shown.

Interrupting and Continuing the Process

You can check, modify or draw zones during processing of a multi-page document by making the program stop when desired. When the process is interrupted, the image of the page being processed will be displayed. You can modify/draw zones and change settings not disabled at this time. After checking, modifying or drawing zones you can re-start the processing of the page or abandon the whole process, leaving the last page unrecognized.

See the topics "Working with Zones" and "Working with Table Zones" in Chapter 4 for more information on zoning.

To preset the program to stop after each image is scanned/loaded:

⇨ Press the *Stop for (re-)zoning* button in the Main toolbar or

⇨ Choose *Stop for (re-)zoning* in the Process menu.

The state of this button can be changed while processing is interrupted.

To stop the process during recognition:

⇨ Click on the *Interrupt* button in the Main toolbar (available during processing only).

To re-start processing:

⇨ Click on the *Continue* button in the Main toolbar (available in interrupted state only) or

⇨ Choose *Continue* in the Process menu.

To abandon processing:

⇨ Click on the *Stop* button in the Main toolbar (available in interrupted state only) or

⇨ Choose *Stop* in the Process menu.

All previous pages will remain, the current page will be unrecognized, but its image will remain. No further pages will be processed.

Recognizing Images in a Document

Some reasons why images in a Recognita Document might need to be (re-)recognized:

- They were originally loaded without recognition, because manual zoning was required.
- The recognition results are not satisfactory because of a wrong setting (for example language, dictionary, brightness, etc.).
- The recognition results are wrong because of improper zone positions/types or incorrect image orientation, etc.

You can (re-)recognize:

- The image on the current page
- All the images
- All unrecognized page images
- Images on selected pages

To recognize the image of the current page:
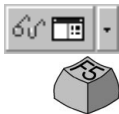
⇨ Make sure the multi-state button on the Editing toolbar shows the image shown here and click on it or

⇨ Choose *Recognize>This Page* in the context menu of the image pane.

To recognize all page images:
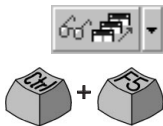
⇨ Make sure the multi-state button on the Editing toolbar shows the image shown here and click on it or

⇨ Choose *Recognize>All Pages* in the context menu of the image pane.

To recognize the unrecognized page images:

⇨ Make sure the multi-state button on the Editing toolbar shows the image shown and click on it or

⇨ Choose *Recognize>Unrecognized Pages* in the context menu of the image pane.

To recognize images on selected pages:

1. Select the page(s) to be recognized from the Browser List.
2. Choose *Recognize Page(s)* in the Browser's context menu.

Working with Documents

After a document has been created, you can further process it in different ways, depending on its contents, your goals and working method. This section gives an overview of the possibilities.

Creating output (see later in this chapter):

- Save or send the document. You can open and work on it later.
- Save or send some or all of the recognized text in a format you choose.
- Save or send some or all of the images in a format you choose.
- Drag-and-drop text and/or graphics to other applications.
- Print text and/or images.

Revising the recognized text (see Chapter 4):

- Check and edit the text manually.
- Start proofing to find and correct problem places in the text.
- Train characters if necessary.

Zoning for (re-)recognition (see Chapter 4):

- Check automatic zones; correct them if necessary.
- Draw zones manually.
- Load zone templates.

Recognizing images (see the previous section in this chapter):

- (Re-)recognize some or all of the images in the document.

Adding new pages (see the first section in this chapter):

- Add new pages to any part of the document.

This guide also presents a separate chapter, "Working with Documents" which details some of these topics. You can also find detailed information on these topics in the online help.

Saving Documents, Text and Images

After a document is created, you can save its text and images and/or save the document as a Recognita Document file. You can also send text, images and Recognita Documents by electronic mail.

This section describes the following procedures:

- Saving and Sending Documents
- Saving and Sending Text
- Using Advanced Settings for Text Output
- Saving and Sending Page Images
- Using Drag-and-drop and the Clipboard

Saving and Sending Documents

Unless you are going to complete your processing very quickly, you should explicitly save your Recognita Document files (also known as RCD files) shortly after creation. Then they are available in later sessions with all their proofing and training facilities. You can also send your RCD files by electronic mail.

To save a document:

1. Click on the Save Recognita Document button in the Main toolbar or choose Save as Recognita Document from the File menu. The first time a document is saved, the Save as Recognita Document dialog box appears. Choose a location and name for your RCD file and click on *Save*.
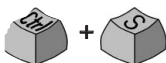
2. Click on the Save Recognita Document button regularly as you work to protect your current changes. The recognized text can be reverted to its last saved state.

To revert text to its last saved state:

1. Select pages from the Browser List whose text you want to revert.
2. Choose *Revert to Saved* from the context menu of the Browser.
3. To revert a whole document, use the command in the File menu.

To send a document as a mail attachment:

⇨ Choose *Send>Recognita Document* from the File menu. Your mail application will be activated with a new empty message containing the document as an attachment.

Saving and Sending Text

After recognition, your Recognita Document contains recognized text. You can save or send it in any of the different output formats supported by Recognita Plus. The formats can be chosen from the Format list of the Save Text(s) and Send Text(s) dialog boxes.

Output formats can be ranked in four groups:

- Text only formats. These include various GWP and ASCII formats, which differ merely in how the original formatting is preserved by line breaks, tabs and spaces.

- Table and spreadsheet compatible formats. Among these you can find tab/comma/quote separated ASCII formats as well as formats for the most popular spreadsheet programs.

- Word processor formats to which Recognita Plus can convert text fully formatted, preserving page layout and including graphics.

- Word processor formats to which Recognita Plus can convert text formatting attributes but maybe not graphics.

Knowing your word processor, DTP or spreadsheet program, you can decide which text format is the most suitable for it.

To save recognized text:

1. Choose *Save Text As* from the File menu or *Save Text* from the context menu of the Browser. The Save Text(s) dialog box appears.
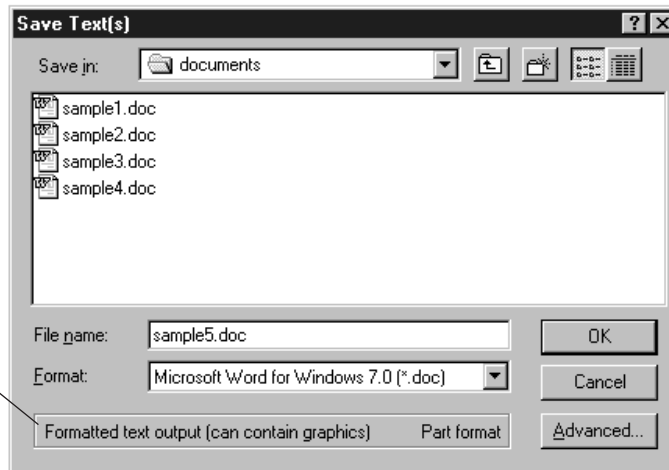
Information on the currently selected format and the current format level

2. Choose an output format from the Format list.

3. Set *Advanced* settings if necessary (see next section).

4. Select folder location, enter file name and click on OK.

To send recognized text as a mail attachment:

1. Choose *Send>Text* from the File menu. The Send Text by Mail dialog box appears.

2. Choose an output format from the Format list.

3. The *Advanced* option is also available for sending (see next section).

4. Choose OK. Your mail application will be activated with a new message containing the text as an attachment.

Using Advanced Settings for Text Output

In addition to choosing a suitable output format, you can have a high degree of control over the way your text document's formatting attributes will be preserved. Click *Advanced* when saving or sending text to display the Advanced Parameters for Saving tabbed dialog box.

Three categories of text format settings

Specify on this tab, which pages to save

Three format levels, choose one of these first

Many of the settings can be automatic or changed manually

Advanced Parameters for Saving

General   Document   Paragraphs   Characters

Format
○ Full   ● Part   ○ Drop

Graphics & Breaks
☑ Retain Graphics   ☑ Retain Page Breaks

Tables
● Cells   ○ Tabs

Margins (inches)
☐ Auto
Left: 1.25   Top: 1.5   Right: 1.25
Bottom: 1.5

Paper Size
☑ Auto   ○ Portrait ○ Landscape

Language
☑ Auto

OK   Cancel   Default   Help

First, you may choose one of three format levels, which correspond to the three main view modes of the built-in editor of Recognita Plus. These format levels are:

- Full format: preserves original page layout; formatted text and graphics are placed in frames.
- Part format: preserves character and paragraph formatting. Text is decolumnized.
- Drop format: preserves text without formatting. Text is decolumnized whenever possible.

Each of these three levels has its own set of remembered settings for document, paragraph and character formatting. Though default values are suitable for most tasks, customizing them may be useful. In full format, many settings are compulsorily Auto, in drop format many are not available. Part format is best for customizing settings.

By default, certain pages are offered for saving and sending. You can select other pages on the General tab.

- If you call saving from the File menu, all pages are offered.
- Using the context menu, the pages selected there will be offered.

You can save each page to a new file by setting the One File per Page option on the General tab.

Saving and Sending Page Images

Scanned images are always embedded in a Recognita Document file; image files can be embedded or simply linked by paths to avoid duplication of the images on your disk. This latter option can be set on the Image tab of the Options dialog box. No matter which is the case, images can be saved to a supported file format. You can also send images by electronic mail.

You can create single or multi-page image files of black-and-white or gray or color images. Images are saved as displayed. The combinations are summed up in a table in the online help.

To save page images:

1. Choose *Save Image As* from the File menu or *Save Image* from the context menu of the Browser.

The Save Image(s) dialog box appears.

Information on the currently selected format
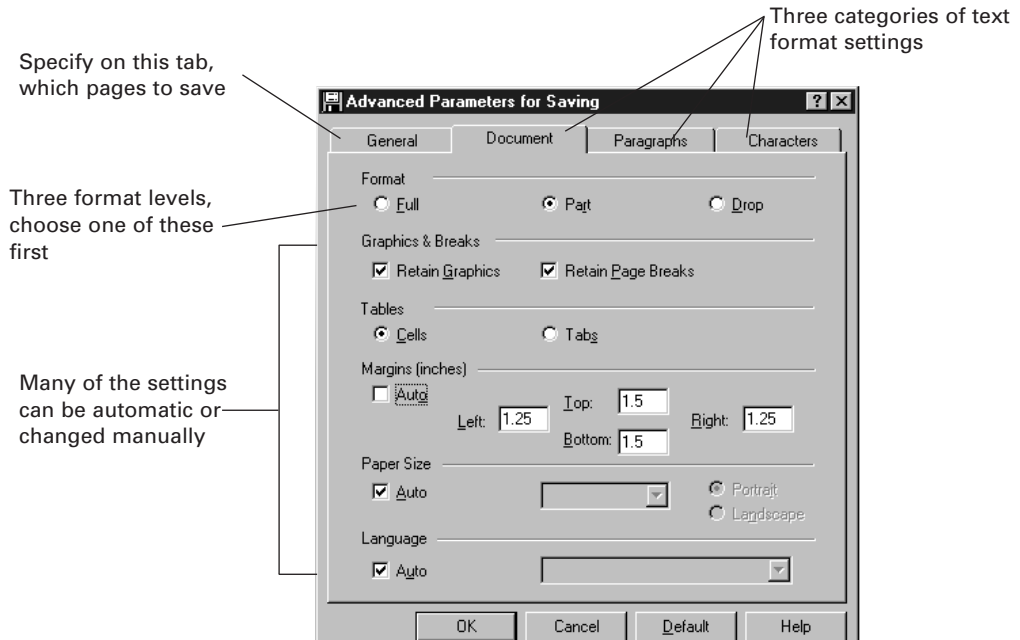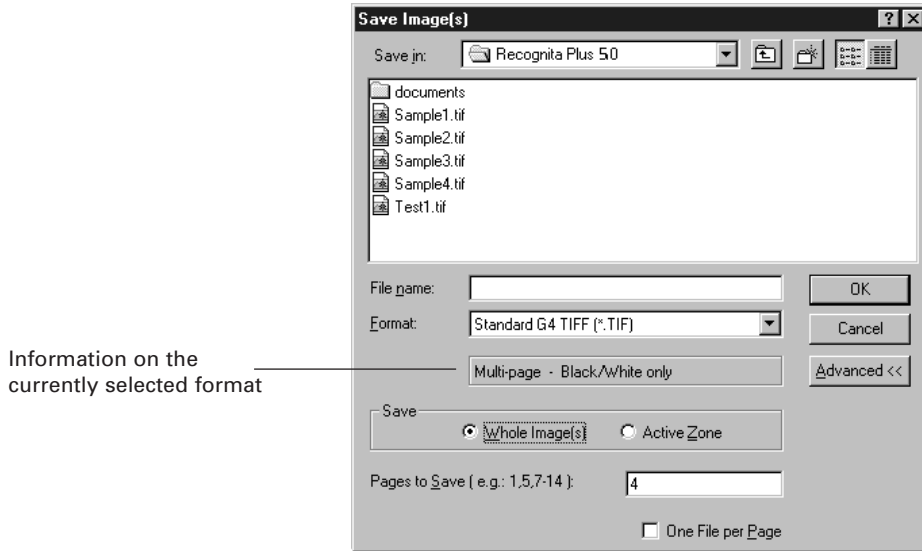
2. Choose an output format from the Format list.
3. If necessary, choose *Advanced>>* to specify the pages to be saved and the One File per Page option.
4. Select folder location, enter file name and click on OK.

To send page images as a mail attachment:

1. Choose *Send>Image* from the File menu. The Send Image by Mail dialog box appears.
2. Choose an output format from the Format list.
3. *Advanced* settings except One File per Page are also available.
4. Choose OK. Your mail application will be activated with a new message containing the image as an attachment.

   If you choose more than one page, they will all be placed in one multi-page image file. If you have chosen a single-page format, you must send each page separately.

By default, certain pages are offered for saving and sending. You can select other pages in both dialog boxes.

- If you call saving from the File menu, the current page is offered.
- Using the context menu, the pages selected there will be offered.

Using Drag-and-drop and the Clipboard

You can select certain parts of a Recognita Document for drag-and-dropping or copying to the Clipboard.

In the text pane you can select the following items for transferring:

- A part of the text, recognized within one zone. Use standard selection methods to select text.

- All the text on the current page. Choose *Select Page* from the context menu of the text pane or *Select Text of Page* from the Edit menu.

- All the text in the document. Choose *Select Text of All Pages* from the Edit menu.

- Graphics in a frame in the text pane. Double-click in a frame containing graphics to select it. You should switch to full format view to do this. (See the section "Editing" in Chapter 4 on the three view modes in the editor.)

In the image pane, you can select any zone by double-clicking in it. The contents of the selected zone will be transferred as image.

Starting Recognition from Other Applications

Recognita Plus can be integrated into your computing environment in various ways. The following methods are provided:

- Direct Connection to Applications
- Recognition Tools in Mail Applications
- Explorer Context Menu Support
- Drag-and-drop from the Explorer.
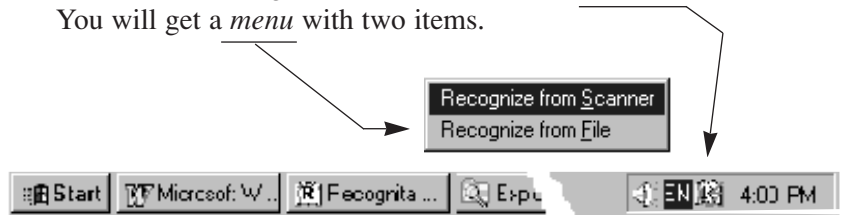
The first two can be enabled during installation or by running the Maintenance Setup of Recognita Plus. The last one is added automatically.

Direct Connection to Applications

This is enabled in Maintenance Setup, and lets you call up Recognita Plus from the taskbar any time you are working in another application. The recognized text will be placed at the cursor position.

To use a direct connection:

1. Start your target application, and place the insertion point at the location where you want the recognized text to be placed.

2. Click on the Recognita Plus direct connection *icon* on the taskbar. You will get a *menu* with two items.

```
┌─────────────────────────┐
│ Recognize from Scanner  │
│ Recognize from File     │
└─────────────────────────┘
```

| Start | Microsof: W... | R Recognita ... | Explo | EN 4:00 PM |

3. Choose *Recognize from File* or *Recognize from Scanner* from this menu. If Recognita Plus is not running it will be started.

4. The recognition process will start according to the menu item selected. The Recognita Plus window will occupy the lower part of the screen.

5. At the end of recognition, text will be placed at the insertion point. The part format level is used for text conversion.

   Right-clicking on the direct connection icon displays a menu to activate the Recognita Plus Options dialog box.

To run recognition in the background:

1. Iconize Recognita Plus after the recognition process is started. The number of the page being recognized will be displayed in the Recognita Plus icon on the taskbar.

2. A flashing icon indicates that recognition is finished. Click on the icon to activate the Recognita Plus window.

3. A message box will be displayed asking you to place the insertion point for text insertion.

4. Only after placing the insertion point should you choose OK in the message box.

   Using background recognition allows you to work on your document while recognition is running. You can even create a new document in the very moment you are prompted to place the insertion point.

Recognition Tools in Mail Applications

You can use Recognita Plus to read image attachments to messages arriving in your mailing system. A new submenu, *Recognita OCR* Tools is added to the menu structure of your mailing system. The following applications are supported:

- Microsoft Exchange
- Microsoft Outlook
- Lotus Notes

You have two basic ways of doing the recognition:

- Reading interactively: this starts Recognita Plus (if necessary); the program recognizes all attachment(s) and places the result in a Recognita Document, ready for proofing and saving.
- Reading non-interactively: This runs in the background, and re-directs the recognition results back into the messaging system as RTF file attachments or as body text, for example for forwarding or replying.

An example of the *Recognita OCR Tools* Menu:

This menu item starts recognition in both cases.

Use Settings to choose between the two ways of doing recognition. Set parameters for non-interactive reading here.
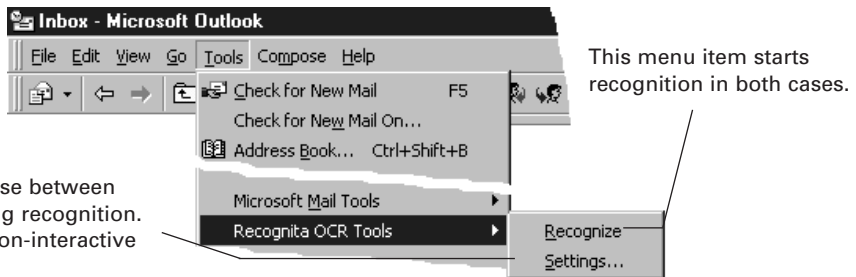
Explorer Context Menu Support

The menu item *Recognize* is added to the context menu of the Explorer (available also on the desktop), if the selected item is an image file of the following types: TIF, BMP, PCX, AWD.

To use the context menu of the Explorer:

1. Select image files in the Explorer or on the desktop.
2. Choose *Recognize*. It starts Recognita Plus (if necessary). Recognita Plus displays the Options dialog box. Change settings if necessary.

3. Click on OK. The recognition starts. Wait for the process to be completed.

4. At the end, the Save Text As dialog box is displayed. Use it to save the recognized text. Recognita Plus remains active.

Drag-and-drop from the Explorer

You can drag-and-drop selected image files onto the icon or the application window of Recognita Plus; it will start, if necessary. The contents of the image files will be recognized just as if they had been opened from inside the program.

Processing and Saving without Display

You can scan or load images, process their contents and save the result so that documents will not be displayed on-screen, but rather saved automatically to one or more output files of the specified type. This method is called *Save without Display*. Typically you will use this for high-volume jobs. You can use it to save images, text or Recognita Documents.

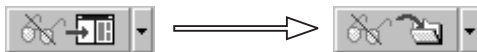### To set the Save without Display mode:

⇨ Choose *Save without Display* from the Process menu. The two possible icons of the Process tool are changed to indicate this special working mode.

The tool with recognition changes as shown:

The tool without recognition changes as shown:

To turn this mode off, click on the menu item again.

To use the Save without Display mode:

1. Set this processing mode as already described.

2. Start processing as you normally would. A dialog box, similar to the Save Text(s) or Save Image(s) dialog appears.

   Important: in this dialog box, you specify saving options! Do not confuse it with the Files(s) to Open dialog box. The latter is displayed after this, if you asked to load image files.

3. Specify location, name and other saving options for your output file(s).

   ⇨ If you start the process with recognition, you can choose a text format or Recognita Document.

   ⇨ If you start the process without recognition, you can choose an image format or Recognita Document.

4. If you want to distribute the incoming pages to more than one file, click *Options* to come up on the Document tab, where you can set the conditions to start a new document, and make other settings.
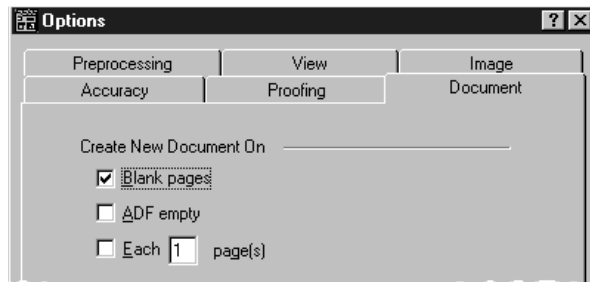
5. Choose OK. If you asked to process image files, the File(s) to Open dialog box will also be displayed. Output files will be generated automatically, according to the specified settings.

   The output files will be given the specified file name plus a four-digit number, starting from 0001 by default. You can enter a different starting number following the file name, enclosed in square brackets. Leading zeros can be omitted. E.g. to start numbering from 200, enter a file name as shown below:

   sample[200]

   The default extension of the chosen file type will be added at saving time.

Working with Documents

Recognita Plus has many features that allow you to further process the documents you created. Which of these possibilities you will use and whether you use them at all depends on you and the complexity of the task you have to accomplish.

In this chapter you will find information on the following topics:

- Working with Zones
- Working with Table Zones
- Correcting the Text
- Navigating in Recognita Documents
- Using the Character Map

Working with Zones

Zones are rectangular areas enclosing printed elements in an image. They identify the parts of the page as text or other elements to be recognized or as graphics to be retained without recognition. Any part of an image outside zones is ignored during recognition. Zones and their reading order are displayed over the images. There is always one and only one active zone on a page; it has handles at each corner and on each side allowing you to re-size it. You activate a zone by clicking inside it.

After scanning or loading an image, the program analyses the page layout, finds text and graphics and creates zones. The program also decides a reading order for the zones.

Zones can also be created manually or by loading a zone template. You can draw new zones or modify the existing ones.

There are six zone types, indicating which recognition engine will run in the zone (typically and most often the Omnifont engine). In zones containing text, distinction is made between flowed text and table zones, also between Language Set (full alphabet) and Numbers Only recognition. All together, these elements form the zone properties.

This section contains the following topics about working with zones:

- Automatic vs. Manual Zoning
- Basics of Manual Zoning
- Basics of Zone Properties
- Basics of Zone Templates.

Automatic vs. Manual Zoning

You may want to disable automatic zoning if you want to recognize only a certain part of your document or the layout of your document is very complicated and you suspect or find that automatic zoning is unsuitable.

⇨ To disable automatic zoning select the setting *Disable Decomposition at Scanning* on the *Preprocessing* tab of the Options dialog box.

Basics of Manual Zoning

Zone handling is available through the toolbar and the context menu of the image pane. This topic describes zoning using the toolbar.

To create zones manually:

⇨ Click in the image to get a crosshair cursor. Drag the mouse to draw a rectangular box.

To resize and move zones:

Catch a handle and drag to resize ——
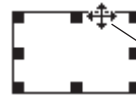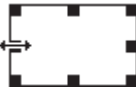
Catch at the border away from handles to move

Toolbar buttons for zoning:

Click this to start reordering zones

Click this to delete all zones ——

Click this to restore original zones
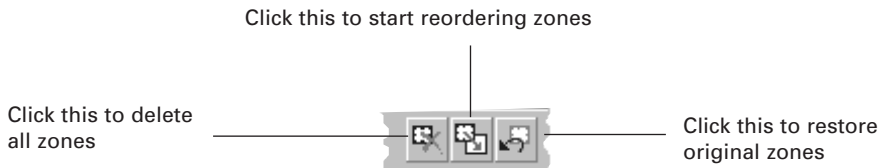
To modify zone order:

1. To start reordering choose the *Reorder Zones* tool or menu item.
2. Click in the last correct zone, then click the zones in the desired reading order. Stop as soon as the order is correct.
3. Click the *Reorder zones* tool again or click outside any zone to finish reordering. Press Esc to abandon reordering.

To delete a single zone:

⇨ Press Del to delete the active zone.

To delete a series of zones:

⇨ Press Ctrl+Del to delete the active zone plus all zones following it.

⇨ Press Ctrl+Shift+Del to delete the active zone plus all zones preceding it.

Basics of Zone Properties

The default settings are suitable for the most common recognition tasks and typically you should not need to change zone properties.

Zones have the following properties:

- One of six recognition engines or graphics
- Text flow: flowed or tabular, only for Omnifont, Dot matrix and Handprinted numbers
- Enabled characters: Numbers Only or Language Set (full alphabet), only for Omnifont, Braille and Dot matrix recognition engines.

The properties are represented by icons and border coloring. To display icons set *Show Properties* in the View tab of the Options dialog box.

Recognition engine ⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯

Enabled characters

Order of zone

Color of border:
- red for flowed text
- blue for tables
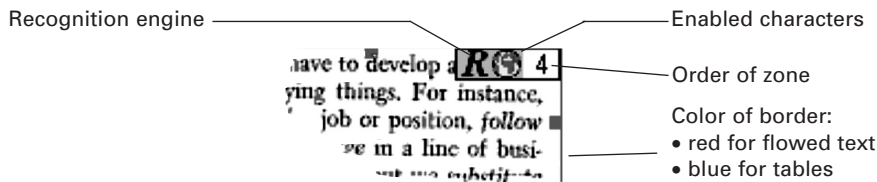
Graphics zones have black borders with gray cross-hatching, without icons.
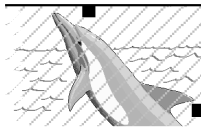
Whether a zone is created automatically or manually, it is first given properties automatically (see later in this section). You can then change any property of an existing zone individually if necessary.

You can set the general zone properties to be applied in future decomposition in the Options toolbar or on the Accuracy tab in the Options dialog box.

Recognition engine
general property ⎯⎯⎯⎯⎯⎯⎯⎯⎯

Enabled characters
general property

To set properties of a zone individually, open the Zone Properties toolbox or use the context menu of the image pane. The active zone's current properties are framed thick. Click a different option to apply it to the active zone.

Recognition engine and graphics property; group of six:
- Omnifont
- Dot matrix
- Handprint (numbers)
- Barcode
- Checkmark
- Graphics

Enabled characters property; group of two:
- Language set
- Numbers only

Text flow property; group of two:
- flowed
- table

How zone properties are set by the program:

- Graphics are always detected automatically.
- Recognition engine:
  ⇨ Decomposed zones take the general setting. If it is set to Automatic, then one of the Omnifont, Dot matrix, Handprint (numbers) or Barcode engines will be chosen, depending on the zone contents detected.
  ⇨ Manual zones inherit the setting of the active zone. When the first zone is drawn it takes the general setting, or Omnifont if Automatic is set.
- Enabled characters (applies for Omnifont and Dot matrix):
  ⇨ Decomposed zones take the general setting.
  ⇨ Manual zones inherit the setting of the active zone.
- Flowing or table text is detected automatically. To disable automatic table detection, press Ctrl and hold down while drawing a zone.
- Braille can be set as the General zone type for a whole document. All existing zones change to Braille zones with red borders (tables are not supported). Manual zone drawing is possible, but all zones in the document must be for Braille; the Language Set/Numbers Only choice remains available. Auto-decomposition places single whole-page Braille zones. Output will be the editable text equivalent of the Braille text. See online help for a list of scanners found suitable for scanning Braille.

Basics of Zone Templates

A zone template file contains information on a set of pre-defined zones (size, location, properties and recognition order) for a single page. Zones can be saved to a template file and loaded whenever needed. You can unload a template, for instance if a wrong one is loaded by mistake.

Zone templates are useful if you want to read many pages or documents with the same page layout. If a template is loaded, automatic decomposition will not be done on new incoming images.

The program will correct a certain level of mis-alignment of template zones which may result from slight displacement of scanned pages.

Right-clicking on the Template field in the Status bar displays a context menu with template-related commands.

To create a template file:

1. Draw or check the zones and set their properties if necessary.
2. Choose *Template>Save* from the File menu or *Save* from the context menu. The Save Template dialog box appears.
3. Enter the name for the template file and choose *Save*.

To load a template file:

1. Choose *Template>Load* from the File menu or *Load template* from the context menu. The Load Template dialog box appears.
2. Select the template file to be loaded and choose *Load*. If a document is open, the Apply template dialog box appears:
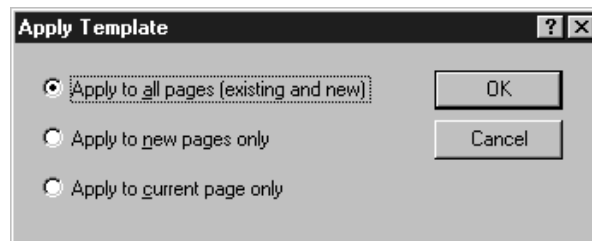
3. Choose one of the three options to apply the template to the desired pages. If the template is loaded on the current or all existing pages, any zones will be removed from them and the template zones will be displayed immediately.

If there is no document open, the loaded template will be applied to new incoming pages.

To unload a template file:

1. Choose *Template>No template* from the File menu or *No template* from the context-menu.

2. If the template is going to be unloaded from an open document, the Remove Template dialog box appears:
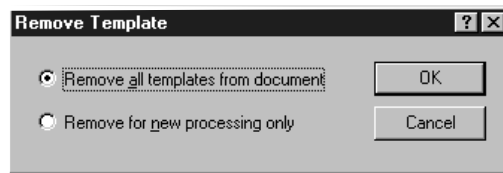
3. Choose one of the two options as desired. If the first one is chosen, all templated zones will be removed from all pages of the document in which no zone editing has been done. To remove a template from the current page only, just edit or delete the zones.

If there is no document open, new incoming pages will be decomposed, if enabled.

Two-page templates are now available for handling two-page forms or books. These templates conserve two zone patterns and apply them to consecutive pages. To save a two-page template, prepare the zones on two consecutive pages, make the first one active, choose Template Save and check the two-page option.

Working with Table Zones

The page layout decomposition automatically distinguishes between flowed text and tables. If a table is detected, the text image is enclosed in a table zone. Tables are also auto-detected when drawing a zone manually unless the Ctrl key is pressed. Tables are indicated by a blue grid over the image.

To toggle between a flowed text (red border) and table zone (blue border), click on the Zone Properties tool midway on the Editing toolbar. This also serves to show the properties of the active zone.

In a table, horizontal gridlines always extend over the full width of the table and can't be shortened. Vertical gridlines do not always extend over the full height of a table:
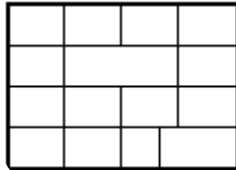
You can edit the gridlines within an active table zone. You do this in the image pane before performing (re-)recognition.

Hints on table editing:

- By default grid snapping is on, making it easier to join vertical lines which do not extend over the full height of the zone. Press Alt or both mouse buttons to enable smooth movement.
- The Ctrl key restricts moving, insertion and deletion of vertical gridlines to the current row.
- If the Ctrl key remains pressed, you can drag the mouse across neighbouring rows to extend insertion and deletion.
- You cannot insert gridlines too close to an existing one. These situations are indicated by prohibiting cursors.

Table editing can be done through the Editing toolbar or the context menu of the image pane. Table zones must be activated before any table gridline editing. You can activate a zone by clicking inside it.

Toolbar buttons for table editing:

Insert rows and columns (vertical and horizontal gridlines)

Delete rows and columns

Delete all rows and columns

To move gridlines:

- You can catch a gridline by the cursor and drag it to a different position.

Gridlines being moved

26.00	25.00	23.00
24.00	22.50	21.50
27.00	25.00	24.00
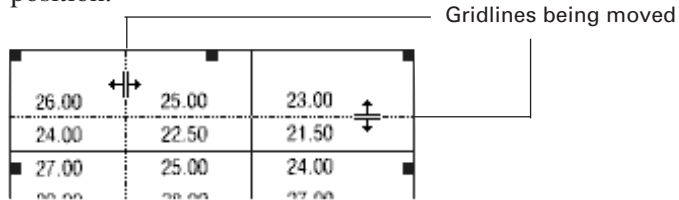
To insert gridlines:

1. Click on the *Insert Columns* or *Insert Rows* tool or use the context menu to get the insertion cursor.

2. Move the insertion cursor to the desired location and click to insert a gridline. Repeat as desired. Press Tab to toggle between the horizontal and vertical cursor.

Cannot insert gridline too close to an existing one

New gridline can be inserted here

26.00	25.00	23.00
24.00	22.50	21.50
27.00	25.00	24.00
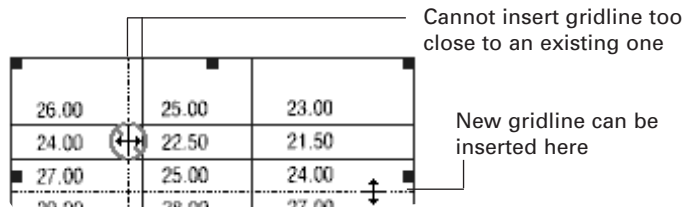
3. To return to a normal cursor, click outside any zone or press Esc.

To delete gridlines:

1. Click on the *Delete Rows/Columns* or use the context menu to get the deletion cursor.

Delete by dragging beyond its neighbor, cursor changes automatically

Delete a gridline by clicking on it

26.00	25.00	23.00
24.00	22.50	21.50
27.00	25.00	24.00
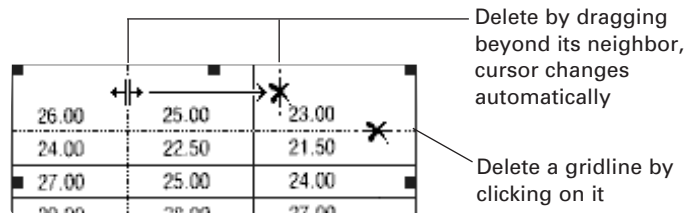
2. To return to a normal cursor, click outside the table zone or press Esc.

To delete all the gridlines:

⇨ Click on the *Delete all Rows and Columns* in the Editing toolbar or use the context menu. This deletes all gridlines in the currently active table zone. The zone preserves its table property; you can then draw your own gridlines.

Correcting the Text

After recognition, the recognized text stored in the Recognita Document is displayed in the text pane. Besides normal recognized characters you may see the following coloured items:

- Suspect characters: characters marked during OCR as unsure appear *highlighted yellow*.
- Non-dictionary words: words not found in the dictionary appear *highlighted green*, provided the main recognition language has a Language Analysis module and it was enabled. The highlight is removed if such a word is changed or stopped on without change during proofing.
- Reject characters: characters the program couldn't identify are represented by *red tildes* (~).
- Missing characters (rare): ones not in the code page selected automatically by Recognita Plus appear in *magenta*. This may happen only if more than one language was enabled and none of your standard Windows code pages can cover all their characters.
- Trained characters: characters changed by training appear in *blue*. They become coloured during training.

To find and correct these, you don't have to rely solely on your eyes; you are also assisted by some tools in Recognita Plus. These are:

- Internal editor complying with standard editing techniques
- Verifiers to compare text and its associated image
- Proofing tool to find problem characters and words
- User dictionaries for proofing
- Training misrecognized characters.

Suspect character and non-dictionary word marking are removed when a word is changed during proofing or typing.

Editing

Recognita Plus comes with an internal WYSIWYG editor having both traditional and OCR-specific features. It is able to display the text and its formatting attributes identified by the OCR engine. It has three main view modes plus a fourth special one. These are:

- Full format: this mode shows the original page layout; both formatted text and graphics are displayed in frames.
- Part format: this mode displays character and paragraph formatting only. Text is displayed decolumnized.
- Drop format: this mode displays the text without formatting.
- Draft mode (rarely used): This mode uses a monospaced font of Recognita Plus for unformatted text display. It can simultaneously display all characters Recognita Plus is capable of recognizing. This may be useful for text display if the fonts required are not installed on your Windows 95 or 98.
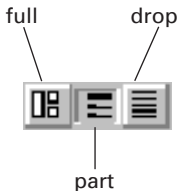
To display the text in full, part or drop format:

full   drop

⇨ Click on the appropriate button at the bottom left of the text pane or

⇨ Choose *Text Format>Full (Part or Drop)* from the View menu.

part

To display the text in draft mode:

⇨ Choose *Draft Mode* from the View menu.

To display the text in different magnifications:

⇨ Choose the desired percentage from the View menu or from the context menu of the text pane.

To make changes to the text:

⇨ Most standard text editing techniques are supported. You can use cut, copy and paste as well as drag-and-drop to edit text.

⇨ Use the Editing toolbar to change character formatting.

⇨ Use the Editing toolbar and the ruler to format paragraphs.

As a rule, you should proof and do any training on the recognized text before doing general editing; the link between text and image may not work on edited characters.

Editing Tables

Once a table zone has been recognized, you can edit both the grid and the contents in the text pane. Recognita Plus respects normal table editing conventions. The following picture contains a summary of cell selection and gridline moving methods:

Select a column by clicking above it

Use the toolbar and the ruler to format text and cells

Catch a gridline to move it

NAME¤	COLOUR¤	SIZE¤	PRICE¤
Item·1¤	green¤	22.50¤	21.50¤
Item·2¤	blue¤	25.00¤	24.00¤
Item·3¤	pink¤	28.00¤	27.00¤
Item·4¤		79.00¤	75.00¤
Item·5¤	black¤	124.00¤	119.00¤
Item·6¤	magenta¤	154.00¤	149.00¤
Item·7¤	white¤	77.00¤	75.00¤
Item·9¤	orange¤	67.00¤	66.00¤
Item·10¤	purple¤	57.00¤	56.00¤

Select a row by clicking before it

Select a single cell by clicking in its left margin

By dragging the mouse you can expand the selection to neighboring rows, columns and cells. Use the context menu of the text pane to do cell editing:

- Use *Split Cells* to split all selected cells in two. This can be used to insert an empty column to the right of a selected column.
- Use *Merge Cells* to merge all selected cells within a row.
- Use *Insert Rows* to insert empty rows before the selected rows. As many rows will be inserted as were selected.
- Use *Delete Rows* to delete the selected rows.

Other editing hints:

- To insert a new row at the bottom of the table, click in the bottom right cell and press Tab.
- To place a tab inside a cell, use Ctrl+Tab.
- Press Del to delete the contents of the selected cells.

Verifiers

Recognita Plus links recognized characters to their original image. Verifiers display these images to make correcting the text easier. Enable or disable the verifiers on the View tab of the Options dialog box. The image pane verifier can be enabled together with either the pop-up or the dynamic verifier.

Pop-up verifier:

- Double-click on a character to be checked in the text pane. The image of the clicked character or space will be centred and shown red in a verifier window. Click anywhere to close the window.

Dynamic verifier:

- Click in text to open this. Its display is the same as the pop-up verifier, but it remains open, tracking the editing position.

Image pane verifier:

- The image of a clicked text pane character is framed blue in the image pane. The image tracks the editing position. Change the image pane magnification to see more or less context.

The picture below shows two verifiers activated:

Image fragment with the clicked character and its neighbors

Clicked word in the editor

Image of clicked character framed blue in image pane
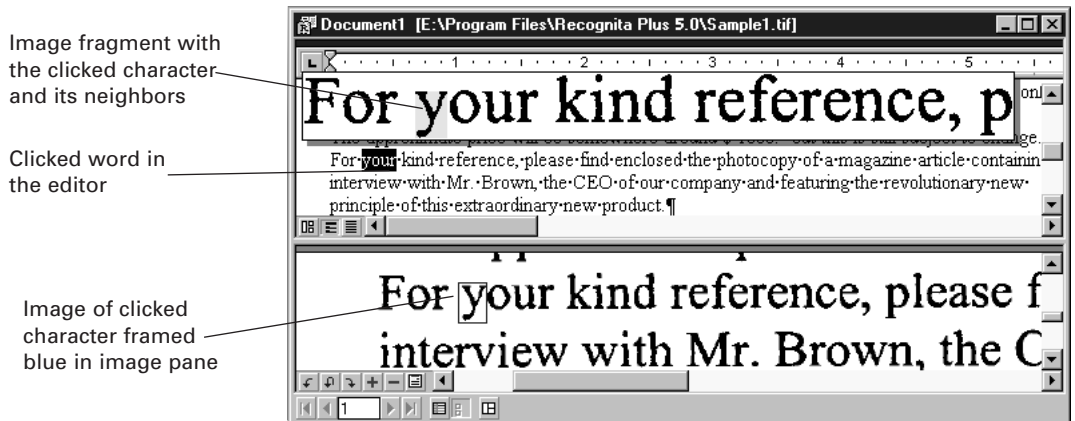
Proofing

Recognita Plus has a special find-and-replace tool for proofing the recognized text. It can help you to find and replace:

- Suspect characters (highlighted yellow)
- Non-dictionary words flagged during recognition (highlighted green)
- Any non-dictionary word found during proofing
- Reject characters (red tilde by default)
- Characters changed by training (blue)
- User defined character strings (e.g. frequently misrecognized character-pairs).

The proofing language and the different stopping conditions can be set on the *Proofing* tab of the Options dialog box. By default, the proofing language is the same as the one used for recognition. It can be changed before or during proofing, for example for different sections of a multi-lingual document.

When the proofing process reaches the end of a page and the next page is loaded, the corrected page will be marked as *proofed*. This is indicated by a checkmark in the Proofed column of the Browser. Once a page is marked as proofed, it will be skipped during future proofing. You can toggle the proofed flag manually by choosing *Toggle Proofed Flag* from the context menu of the Browser. If the proofed flag is turned on again, the suspect character and non-dictionary markings are displayed and you can proof the page again.

To start (and also to stop) proofing:

⇨ Click on the *Proof* tool in the Main toolbar or choose *Proof* from the Edit menu.

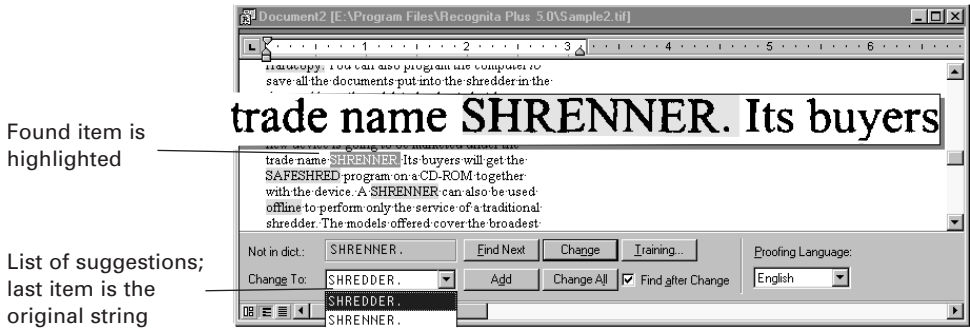The proofing dialog bar appears at the bottom of the text pane.

To find a problem place in the text:

⇨ Click on *Find Next* in the proofing dialog bar.

If enabled, the verifiers are automatically activated on found items.

Found item is highlighted

List of suggestions; last item is the original string

To correct found words using the proofing dialog bar:

⇨ Select a suggestion from the dropdown list and click on *Change* (this option is available if a proofing language is selected) or

⇨ Enter your correction in the Change To field and click on *Change* or

⇨ Click on *Add* to add the selected suggestion or corrected word to the user dictionary. For more information on user dictionaries see the topic "User Dictionaries" later in this chapter.

⇨ Click on *Training* to train the found item. For details on training, see the topic "Training characters" later in this chapter.

You can choose *Change All* instead of *Change* to replace all occurrences of a non-dictionary word or a string with its correction, throughout your whole document.

To correct found words using the editor:

⇨ Press Esc when a word is found to move the insertion point from the proofing bar to the word in the text pane for editing there. Press Esc twice if the dropdown list with suggestions is open.

To get suggestions on any word in the editor:

1. Start proofing.
2. Select the word in the editor on which you want to get suggestions. The *Add* button will change to *Suggest*.
3. Click on *Suggest* to get suggestions.
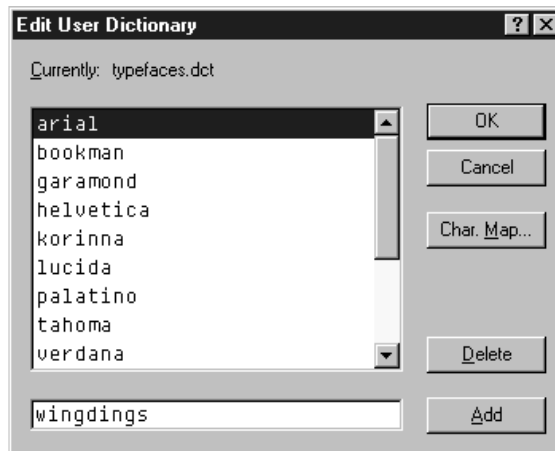
User Dictionaries

In addition to the main dictionary that can be used by the recognition and proofing processes, you can create user dictionaries by adding words during proofing.

User dictionaries can be saved for future use and one can be loaded per document whenever needed. If no user dictionary is loaded, words added will be stored in memory until saved. Loaded and new dictionary information will be used by both recognition and proofing. The name and status of the currently loaded user dictionary is displayed in the status bar.

Right-clicking on the User dictionary field in the Status bar displays a context menu with dictionary-related commands.

To edit a user dictionary:

1. Load the user dictionary. (You can also edit dictionary words added during proofing and not yet saved.)
2. Choose *Edit User Dictionary* from the Edit Menu. The Edit User Dictionary dialog box appears.

3. To add a word, enter it in the textbox at the bottom and click on *Add*.
4. To delete a word, select it from the list and click on *Delete*.

To add words to the user dictionary during proofing:

⇨ Choose *Add* in the proofing dialog bar as described in the previous topic.

To save/load/unload a user dictionary:

⇨ Choose the appropriate menu item from the context menu or from the *User Dictionary* submenu in the File menu.

⇨ User dictionaries can also be loaded and unloaded by clicking on the button with three dots (**'…'**) on the Accuracy tab in the Options dialog box.

Training

Training is the process of associating character shapes (images) with the characters they represent. It can be done *after* recognition on Omnifont or Dot matrix characters.

Most characters need not be trained. As a rule, you should train character shapes which are repeatedly misrecognized or unrecognized. In other words: do not train individual errors caused by accidental spots on the image. You can also train uncommon characters and symbols.

Training can be saved to training files for future use and loaded whenever needed. If no training file is loaded, all new training information will be stored in memory until saved. Loaded and new training will be used by recognition. You can unload training if it is not needed any more in the current document. Reviewing and editing of training files is also possible. The name and status of the currently loaded training file is displayed in the status bar.

When a character shape is trained, the program does the following:

• Corrects the occurrence of the character used for training.

• Looks further down on the same page and checks if the shape of any recognized character is similar to that of the trained one.

• Presents proposed changes to the user for confirmation.

• Corrects all occurrences of the similar characters if confirmed.

Hints for training:

- Always start training at the beginning of the document.
- Use only a few pages to train characters. Training increases recognition accuracy on subsequently added and recognized ones.
- Use separate training files for different types of documents.
- Even if you don't want to save your training, it can be useful to speed up proofing.

Right-clicking on the Training file field in the Status bar displays a context menu with training file related commands.

To train characters:

1. To initiate training you have the following choices:

    ⇨ Click on *Train* in the proofing dialog bar, if you want to train the found item or

    ⇨ Right-click on the character or selected word in the editor and choose *Train* from the context menu. The Training dialog box appears:

Image of character to be trained and its context

Image of character (basic shape) to be trained, colored blue. Unwanted fragments can be detached by clicking.
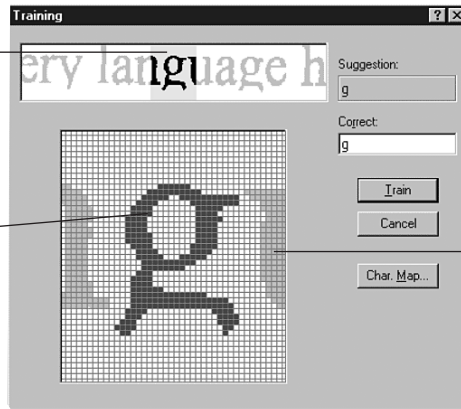
Neighbors, colored yellow. Click to join to the basic shape (blue part).

2. Enter the correct character in the textbox and click on *Train*.

3. If characters with similar shapes are found on the same page, the
    Check Training dialog appears.

- List of words with similar shapes
- Proposed changes in each word appear blue on screen
- Original image of selected word
- Textbox with the selected word
- Context

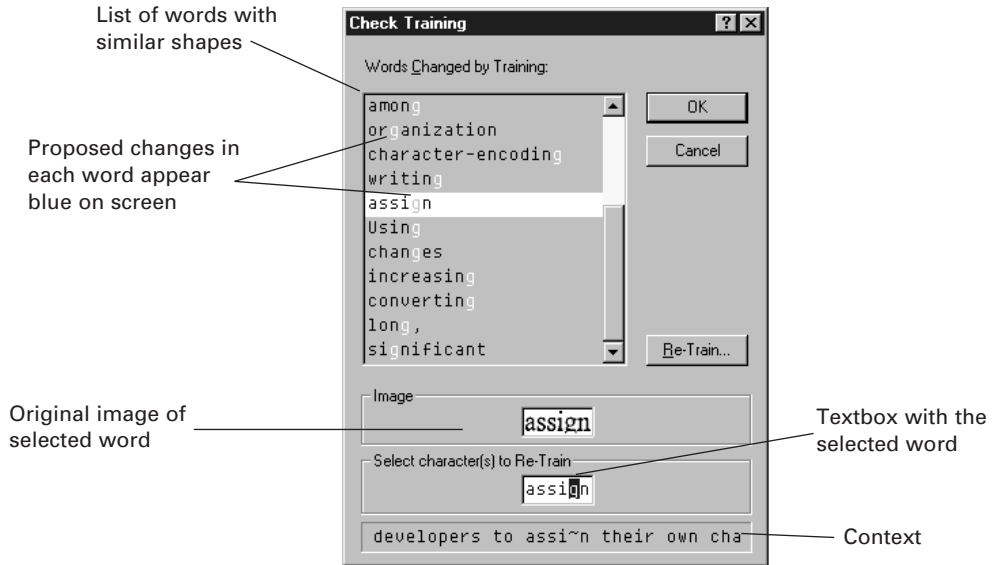4. Check if all proposed changes are correct. Some might be incorrect
    due to the similar shapes of different letters. (For example: 'b' and
    'h', 'q' and 'g', etc.) You have the following choices:

    ⇨ If all words are correct, click on OK. The proposed changes
      will be made; the changed characters will appear blue.

    ⇨ If only a few proposals are incorrect, select an incorrect word
      and click on *Re-train*. The Training dialog box appears where
      you can re-train that single occurrence. Repeat the step as
      required, click on OK when finished.

    ⇨ If many proposals are incorrect, you should choose neither OK
      nor Re-train but *Cancel*. That training will be abandoned.

To save/load/unload training:

⇨ Choose the appropriate menu item from the context menu or from
  the *Training* submenu in the File menu.

⇨ Training files can also be loaded and unloaded by clicking on the
  button with three dots (**'…'**) on the Accuracy tab in the Options
  dialog box.

To review/edit training:

1. Load the training file. (You can also edit unsaved training.)

2. Choose *Edit Training File* from the Edit Menu. The Edit Training File dialog box appears.

3. The trained shapes and associated characters will be displayed. You can enter new characters for a shape and delete unwanted ones.

Navigating in Recognita Documents

Recognita Plus displays one page of a Recognita Document at a time; it is called the current page. You can use the Page Browser to change pages sequentially or randomly. Other tools help you to find pages of the document. Concise information on pages can also be displayed for easy navigation. You can easily copy or move pages within a document or between documents.

This section describes how to work with multi-page documents. The following topics are included:

- Changing Pages
- Using the Browser
- Finding Pages and Text

Changing Pages

To change pages you can use the buttons at the bottom left of each document window. To use the keyboard see the online help.

Go to previous page   Go to next page

Go to first page   Go to last page
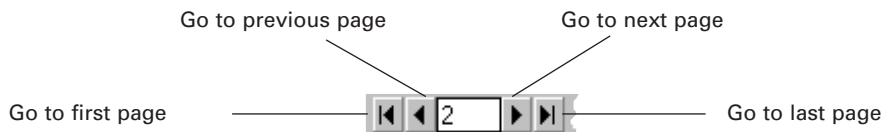
The textbox in the middle shows the page number of the current page. Enter a new page number in it and press Enter to go to the desired page. Press Esc instead if you change your mind.

Using the Browser

The Browser occupies the left or bottom pane in the Recognita Document window. It consists of two parts. The left part displays thumbnail size images of the pages. The right part contains the Browser List with lines, each representing one page.

You can use the Browser for many different things. In addition to displaying information on pages, you can use its context menu to initiate commands. Most of these commands apply to selected page(s), for instance (re-)recognizing, opening and deleting pages, saving text and images, finding text, etc.

Browser list with customizable columns. To customize, choose *Columns* from the View menu

Last line shows statistics summary

Browser page images

Click here to show or hide the Browser pane

Click here to show or hide the Browser page images

Click here to change between vertical and horizontal splitting
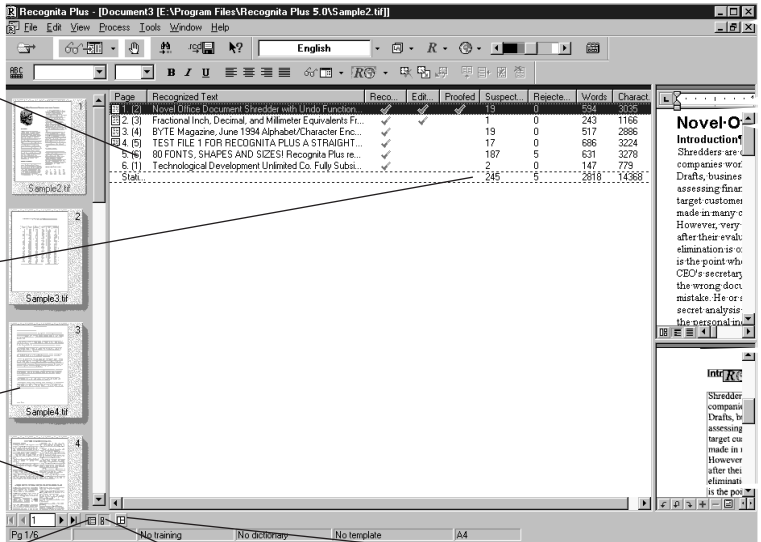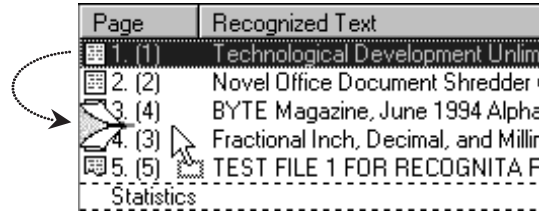
This topic describes the following features of the Browser:

- Moving and copying pages: useful if the order of pages is wrong.
  - ⇨ Move pages to a different location within the same document.
  - ⇨ Move or copy pages to another document.
- Quick display of the recognized text: you can use it to quickly see the results of recognition, without changing pages.
- Adding notes to pages: later you can find pages containing given keywords in their notes column.

To move/copy pages:

1. Select the pages to be copied/moved from the Browser List. Standard selection methods can be applied.

2. Click on a selected item and hold down the mouse button. Drag the pages to move them to the target location in the same or in another document. To copy pages to another document, hold down the Ctrl key when releasing the mouse. The target location is continuously indicated by an icon as shown.

Selected page is moved; the opening pages icon indicates the current target location

3. Release the mouse at the desired location.

To quickly display the Recognized Text:

1. Customize the Browser's columns so that the Recognized Text column is selected.

2. Activate the Page Browser pane by clicking on any part of it.

3. Move the cursor onto the Recognized Text column then leave it there. The recognized text (as much as possible) appears in a popup window just like a ToolTip. By moving from line to line, you can easily and quickly view the text of many pages.
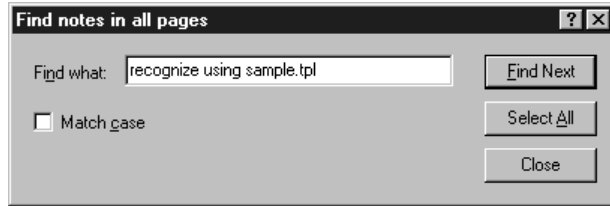
To add notes to Pages:

1. Customize the Browser's columns so that the Note column is selected.

2. You have two choices:

   ⇨ Select the line of the desired page and click on the Note field (or press F2) to key in note text in-place.

   ⇨ Choose *Edit Note* from the context menu of the Browser to display the Edit Note dialog box.

Finding Pages and Text

You can search for pages containing a certain string in their note field, or find strings in recognized text quickly without opening the pages. If pages are selected in the Browser and searching is started from its context menu, only the selected pages will searched. Otherwise all pages are searched.

To find pages with a given string in their note field:

1. Choose *Find>In Notes* from the Edit menu or from the context menu of the Browser. The Find notes … dialog box appears.

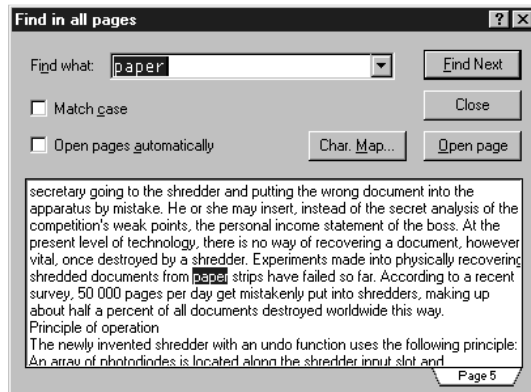|		
Find notes in all pages		**? ✕**
Fi<u>n</u>d what:	recognize using sample.tpl	<u>F</u>ind Next
☐ Match <u>c</u>ase		Select A<u>l</u>l
		Close

2. Enter a string and click on Find Next. The first page whose note field contains the given string will be selected in the Browser. Repeat as desired to find further occurrences.

3. Click Select All to select and highlight all the pages containing the string in their note (for instance ready to be drag-and-dropped to a new location).

To find strings quickly in the recognized text:

1. Choose *Find>In Text* from the Edit menu or from the context menu of the Browser. The Find in … pages dialog box appears.

Find in all pages		**? ✕**
Fi<u>n</u>d what:	paper ▾	<u>F</u>ind Next
☐ Match <u>c</u>ase		Close
☐ Open pages <u>a</u>utomatically	Char. Map...	<u>O</u>pen page

secretary going to the shredder and putting the wrong document into the apparatus by mistake. He or she may insert, instead of the secret analysis of the competition's weak points, the personal income statement of the boss. At the present level of technology, there is no way of recovering a document, however vital, once destroyed by a shredder. Experiments made into physically recovering shredded documents from paper strips have failed so far. According to a recent survey, 50 000 pages per day get mistakenly put into shredders, making up about half a percent of all documents destroyed worldwide this way.
Principle of operation
The newly invented shredder with an undo function uses the following principle: An array of photodiodes is located along the shredder input slot and

Page 5

2. Enter a string and click on Find Next. If found, the program displays the text containing the string, which will be highlighted. If the text found is on the current page, the editor also highlights the occurrence. Pages containing the searched text will be opened only if you click on *Open Page* or the *Open pages automatically* option is set.

Using the Character Map

The Character Map is a small popup window displaying a table of characters. It has two forms:

- Displaying all 464 characters Recognita Plus can recognize. Usually, you see this table. Use it to insert characters in the text pane or into certain textboxes in dialog boxes. It is useful especially if you want to insert special symbols or non-keyboard characters, for example for training, searching, etc.

- Displaying the characters and code values of the selected code page. This is available on the Character tab of the Advanced Parameters for Saving dialog box, and is displayed only for information. The codes help you create a user-defined code page.

To display the Character Map:

⇨ Click on the Character Map tool in the Editing toolbar or choose *Char. Map* in dialog boxes where available.

Characters of the current recognition language(s) displayed black

Other characters displayed gray

Characters enabled individually on the Accuracy tab displayed blue

To insert a character from the Character Map:

1. Place the insertion point to the desired location (in the text pane or in a textbox of a dialog box).

2. Click on a character to insert it. Any character can be inserted, regardless of its current status (color).

Improving Recognition Accuracy

Successful recognition depends mainly on two things: the quality of your document and the current settings of Recognita Plus. This chapter gives you some practical advice and tells which settings are the most important to achieve the highest accuracy possible. But don't forget: your practical experience is at least as important as proper settings; even the best software cannot do without it.

In this chapter you will find information on the following topics:

- Scanner Settings
- Languages and Language Analysis
- Accuracy Troubleshooting

Scanner Settings

Scanner settings can be set on the Scanner tab of the Options dialog box or on the TWAIN data source's own user interface, if the TWAIN Basic driver was chosen at installation or setup. Brightness can also be set on the Options toolbar. Available settings may vary for different scanner models. The most important of them are:

- Brightness
- Resolution
- Scanning Mode.

Setting Correct Brightness

A proper brightness setting results in characters whose contours are neither broken, nor run into each other. For good quality documents the default value (usually 50%) gives good results. You can examine the quality of the scanned image in the image pane at the maximum zoom setting.

Sample images scanned with different brightness values:

extensive research — Unsuitable

extensive research — Unsuitable

extensive research — Tolerable

extensive research — Good

extensive research — Good

extensive research — Best

extensive research — Best

extensive research — Good

extensive research — Good

extensive research — Tolerable

extensive research — Unsuitable

You should of course try to reach the optimum image quality, however this is not always possible. Recognita Plus tolerates broken lines and touching characters up to a certain degree, so you shouldn't worry too much.

You may still get reasonable accuracy, even if the image quality is poor, by enabling Language Analysis. See the next section in this chapter for details on language-related settings.

Setting Proper Resolution

A proper resolution is also necessary to get good results. Though document quality must also be considered, as a rule, you should set

- 300 dpi for letters larger than 8 points,
- 400 dpi for letters smaller than or equal to 8 points.

Do not set resolution higher than 400 dpi; it may cause more harm than good to your results.

Check the resolution of image files. To display resolution, enable the Resolution column in the Browser List.

Choosing Proper Scanning Mode

Recognita Plus has four basic scanning modes you may choose from:

- Scan B/W: most often used. Scans black and white with the given brightness setting. Choose this for documents of reasonable or good quality.
- Scan B/W with Auto-brightness: scans in gray using either your scanner's image optimization software or Recognita's own facility to derive an optimum black and white image. Gray image is not retained. Choose this setting only on poor quality documents where the contrast varies on a single page or from page to page.
- Scan Gray: scans in gray. This mode is used primarily to have grayscale images embedded in a Recognita Document for display and exporting. It also derives an optimum black and white image giving similar results to that of the Scan B/W with Auto-brightness mode.

- Scan Color: This is offered if your scanner supports color. It allows color images to be displayed in the image pane and Browser. These color images can be printed, sent or saved to image files. With color scanning, graphics zones in text files will also be displayed in color whenever Full or Part Format it set. They can be printed in color. They can also be sent or saved in color, provided Retain Graphics is set and a suitable output format selected.

Languages and Language Analysis

In this section you will have information on the following topics:

- Recognition Languages
- Language Analysis (using dictionaries)
- Omnifont Recognition Methods

Recognition Languages

It is very important to define the set of characters, to be enabled for recognition. Typically and most often you do this simply by setting the language of your input document. This has a major impact on how its characters will be recognized. Recognition languages can be set in the dropdown list on the Options toolbar or on the Accuracy tab of the Options dialog box.

Click here to display languages

Main recognition language appears bold on the toolbar

English, German

When you set a language, you add the characters needed for that language to the set of characters enabled for recognition. This is called the Language Set. Selecting more than one language for a multi-lingual document extends this set. Punctuation characters, numbers and other common symbols are always enabled.

If more than one language is selected, the first one selected is set as the main recognition language. Click with Ctrl on a different language to make it the main one. Language Analysis can be enabled if the main recognition language has a dictionary. The main recognition language is also used to find a suitable code page for exporting text.

The predefined set of digits, called Numbers Only, is an alternative to the Language Set if your document contains only or almost only numbers. Both settings can be extended by enabling additional characters individually.

To see which characters are enabled for each language, see the topic "Languages and Accented Letters" in the online help. Its Language section also gives advice on handling multi-lingual documents, and Code Pages for text export.

How to Customize the Language List

This feature is new to Recognita Plus 5.0. On delivery, the toolbar language list displays the seventeen languages for which Language Analysts are available. To customize this list:

1. Go to the Accuracy panel of the Options dialog box.
2. Click Customize Languages... for an alphabetical listing of all 114 supported languages.
3. Shorten the list if desired by de-selecting some continents or categories.
4. Select languages in either list and use the Add and Remove buttons to keep just the languages you need in the toolbar list. Added languages appear at the bottom of the list.
5. To reorder the list, select one language at a time and use the up or down arrow buttons.

Language Analysis (using Dictionaries)

Language Analysis includes the process of using dictionaries during recognition. If it is enabled, the Omnifont OCR engine consults the dictionary of the main recognition language and also the current user dictionary – if one is loaded – during recognition to verify and correct words being recognized, thus increasing accuracy. It is also used to mark any non-dictionary words in the text recognized by the Omnifont or Dot matrix engines.

See the next topic "Omnifont Recognition Methods" on using Language Analysis.

Omnifont Recognition Methods

The Omnifont recognition engine has six accuracy/speed levels, often called recognition or OCR methods. These levels define how many recognition passes are applied on each page and whether Language Analysis is used or not.

This is the main recognition language

The icon ᴸᴀ means Language Analysis can be used for this language

These languages are also selected

Slider to set one of the six accuracy/speed levels; enable Language Analysis here

Add individual characters to the Numbers Only set here

Add individual characters to the Language Set here

The six recognition methods:

Recognition methods can be set on the dropdown tool in the Options toolbar or on the Accuracy tab of the Options dialog box.

Fastest

Most accurate

- Level 1: One-step reading without Language Analysis. This is the fastest. Use it:
  - ⇨ For very good quality documents.
  - ⇨ When speed is more important than accuracy.

- **Level 2: One-step reading, with Language Analysis. Use it:**
  - ⇨ For good quality documents containing typical language.
  - ⇨ For pages containing very little text.
  - ⇨ When speed is rather more important than accuracy.

- **Level 3: Two-step reading without Language Analysis. Called Balanced. Use it for:**
  - ⇨ Typical documents of reasonable quality.
  - ⇨ Reading languages for which Language Analysis is not available.
  - ⇨ Documents with many proper nouns or non-dictionary words.
  - ⇨ Pages with two or more languages.

- **Level 4: Two-step reading with Language Analysis. Use it for:**
  - ⇨ Typical documents of reasonable quality.
  - ⇨ Documents without too many proper nouns and non-dictionary words.

- **Level 5: Three-step reading with Language Analysis. Most accurate with single-engine recognition. Use it for:**
  - ⇨ Degraded documents.
  - ⇨ Languages with Language Analysis but not supported by the second engine: Catalan, Czech, Greek, Hungarian, Polish and Russian.

- **Level 6: Most accurate three-step reading with Language Analysis and dual-engine recognition. Use it:**
  - ⇨ When maximum accuracy is vital and slower processing does not matter.
  - ⇨ On powerful fast computers.
  - ⇨ For the languages supported by both engines: Danish, Dutch, English, Finnish, French, German, Italian, Norwegian, Portuguese, Spanish or Swedish.

Accuracy Troubleshooting

Many things can cause unexpectedly poor or incorrect recognition results. This section summarizes typical problems and their reasons. A new setting and re-recognizing can solve most of the problems.

Poor recognition:

- Low image quality due to poor brightness or resolution setting.

Many green highlights, though words are mainly correct:

- Wrong language was set with Language Analysis enabled.
- Wrong or missing user dictionary.

Many wrong characters, earlier recognized correctly:

- Wrong or missing Training file.
- Characters earlier enabled individually are no longer set.

Missing accents or misrecognized accented characters:

- Wrong language was set.

Many nonsense words and garbage characters:

- Wrong recognition engine was set, (for example, Dot matrix recognition engine was used instead of Omnifont or vice versa) or automatic recognition type detection did not work correctly.
- Image orientation is wrong. Either you placed the document in the scanner the wrong way or the automatic orientation detection was incorrect. You can rotate and (re-)recognize the images.

Garbage characters in certain lines:

- Improperly placed zones cutting a line in two parts.
- Improperly placed gridlines in a table zone.
- Improper margin setting in Tools/Options/Area.

Text contains mainly numbers and reject symbols:

- Omnifont, Braille or Dot matrix recognition was run on zones containing normal text but with the Numbers Only property set, or the Handprinted numbers recognition engine was used on normal printed text.

We trust Recognita Plus will accompany and serve you well, on the road to greater recognition.

Don't forget the different sources of help you can turn to:

- This User's Guide
- Online help
- Homepage: www.caere.com/recognita
- Fax: (36 1) 452-3710
- Tel.: (36 1) 452-3706

We suggest you enter your registration number below as soon as you receive it. Then, you will have it on hand if you need to call on product support.

Registration number:

The information contained in this document is subject to change without prior notice.