



Readiris™ Pro 12

User Guide

Table of Contents

Copyrights	1
Chapter 1 Introducing Readiris	3
Save time, avoid retyping.....	3
The Readiris series	6
Chapter 2 Installing Readiris	9
System requirements	9
Software installation	10
Uninstalling the software	11
Software registration	11
Product support	12
Chapter 3 Getting started.....	13
Running Readiris	13
Using the OCR Wizard	13
User interface.....	15
Changing the user interface language	17
Chapter 4 The Readiris SmartTasks	19
Using the Readiris SmartTasks	19
Chapter 5 Scanning documents.....	23
Selecting the options	23
Opening image files	24

Scanning paper documents.....	26
Chapter 6 Adjusting scanned documents.....	33
Chapter 7 Saving documents as image files	39
Chapter 8 Windowing documents.....	41
Windowing documents automatically	41
Windowing documents manually.....	43
Using windowing templates.....	47
Chapter 9 Recognizing documents	51
Introduction.....	51
Selecting the document language.....	52
Defining the document characteristics	55
Using interactive learning	57
Using font dictionaries	59
Chapter 10 Formatting and saving documents.....	61
Formatting documents	61
Formatting text documents.....	63
Formatting table-based documents	67
Creating PDF documents	71
Selecting the PDF options.....	72
iHQC compressing PDF documents	73
Repurposing PDF documents.....	75

Creating XPS documents	75
Selecting the XPS options.....	77
iHQC compressing XPS documents	78
Selecting the graphics options.....	79
Chapter 11 Saving and loading settings	81
Chapter 12 Recognizing multipage documents.....	83
Opening and recognizing multiple image files.....	83
Scanning and recognizing multipage documents	85
Editing multipage documents.....	86
Chapter 13 Recognizing handprinted text	89
Chapter 14 Recognizing barcodes.....	93
Index	95

Copyrights

ReadirisPro12-dgi-080610-06

Copyrights © 1987-2010 I.R.I.S. All Rights Reserved.

I.R.I.S. owns the copyrights to the Readiris software, to the online help system and to this publication.

The information contained in this document is the property of I.R.I.S. Its content is subject to change without notice and does not represent a commitment on the part of I.R.I.S. The software described in this document is furnished under a license agreement which states the terms of use of this product. The software may be used or copied only in accordance with the terms of that agreement. No part of this publication may be reproduced, transmitted, stored in a retrieval system, or translated into another language without the prior written consent of I.R.I.S.

This user guide utilizes fictitious names for purposes of demonstration; references to actual persons, companies or organizations are strictly coincidental.

Trademarks

The Readiris logo and Readiris are trademarks of Image Recognition
Integrated Systems S.A.

OCR, ICR and barcode technology by I.R.I.S.

AutoFormat and Linguistic technology by I.R.I.S.

BCR and field analysis technology by I.R.I.S.

iHQC compression technology by I.R.I.S.

XML parser developed by Apache. This product includes software developed
by the Apache Software Foundation.

All other products mentioned in this user guide are trademarks or registered
trademarks of their respective owners.

CHAPTER 1

INTRODUCING READIRIS

SAVE TIME, AVOID RETYPING

Congratulations on acquiring Readiris. This software package will undoubtedly be of great help in recapturing your texts, tables, graphics, barcodes and handprinted texts.

As efficient as computers are, you have to key in your information first. If you have ever retyped a 15 page report or a large table of figures, you know how tedious and time-consuming it can be. Use this state-of-the-art OCR package to automatically convert paper documents or scanned image files into text searchable and editable documents that can be archived and shared. Two recognition modes are available: one ensures maximal speed, the other guarantees optimal OCR accuracy.

Scan a printed or typed document, indicate the zones you want to recognize with Readiris - or have the system detect them for you - execute the character recognition and export the document to your word processor. Documents composed of many pages are processed from start to finish in a single effort. A few mouse clicks beat long hours of work as Readiris converts your paper documents into editable computer files: it's up to 40 times faster than manual retyping.

The wizard smoothly guides you through the settings required to operate Readiris, allowing you to obtain quick and easy results. Or use the SmartTasks to speed up the process even more. You can send the reading results directly to your word processor or

spreadsheet, archive them as PDF or XPS files, etc. To recognize faxes and convert PDF documents, drag their image files from Windows Explorer to the Readiris application window. Or send an image promptly to Readiris via the context menu.

Readiris recognizes tabular data and recreates them as worksheets in your spreadsheet software or as table objects inside your word processor; your numeric data are immediately ready for further processing.

Readiris is based on the most advanced recognition technologies. Font-independent text recognition is complemented by self-learning techniques. The system is able to learn new characters and words through contextual and linguistic analysis. This means that the OCR accuracy of the recognition system will improve as it goes along.

Readiris supports up to 128 languages: all American and European languages are supported, including the Central-European, Baltic and Cyrillic languages as well as Greek and Turkish. Optionally, Readiris can read Arabic, Farsi and Hebrew documents and four Asian languages - Japanese, Simplified and Traditional Chinese and Korean. Readiris even copes with mixed alphabets: the software detects “Western” words that occur in Greek, Cyrillic, Arabic, Hebrew and Asian documents - many untranscribable proper names, brand names, etc. are written using the Western symbols.

Readiris uses linguistics *during* the recognition phase, not afterwards. As a result, Readiris recognizes all kinds of documents with top accuracy, including low-quality documents, faxes and dot matrix printouts. It copes beautifully with badly scanned and copied documents containing too light or dark font shapes. Joined characters are resolved while fragmented characters, such as dot matrix symbols, are recomposed.

Besides that, Readiris has an (optional) user verification function. When activated, the user verification function ("Interactive learning") not only flags the characters the recognition system isn't sure of but also allows to increase the system's accuracy. All

solutions you confirm are memorized, increasing the system speed and confidence and rendering the system more intelligent as you go along. This powerful learning tool also allows you to train Readiris on special characters such as mathematical symbols and dingbats and to handle distorted fonts.

To increase your productivity further, Readiris not only recognizes your texts, but can format them for you as well. Various levels of formatting are available. When you make use of “autoformatting”, Readiris recreates a facsimile copy of the scanned document: the word, paragraph and page formatting of the original document are retained. Similar typefaces are used, the point sizes and type styles as used in the source document are maintained across the recognition. The placement of columns, text blocks and graphics follows your original documents. Readiris can even include the background photo of a scanned page in the recognized document. And as Readiris supports grayscale and color scanning effortlessly, you can recapture any graphics - be they line art, black-and-white photos or color illustrations. When a document contains tables, Readiris reorganizes them in real cells and recreates the cell borders of the original tables.

In other words, Readiris allows you to archive a true copy of your documents, be it editable and compact text files instead of scanned images.

Barcodes that occur on a scanned page can also be read, and the same goes for handprinted text, provided you write well-spaced “block letters”.

Readiris is Twain compliant and supports a wide range of flatbed and sheetfed scanners, “all-in-one” devices or “MFPs” (“multifunctional peripherals”) and digital cameras. Interval scanning allows you to scan multipage documents efficiently when your scanner is not equipped with a document feeder.

THE READIRIS SERIES

The table below gives an overview of the available versions:

Readiris Home 12	
Limited features	
25 recognition languages	
Supports PDF, DCX, DJV, DJVU, JPG, JPEG, J2C, J2K, JP2, PNG, TIF, TIFF, BMP, PCX	
Generates PDF Image-Text, DOCX, ODT, WordML, SpreadsheetML, RTF, HTM, XML, TXT, TIFF, etc. output	
Readiris Pro 12	Readiris Corporate 12
Basic features	Basic features
128 recognition languages	128 recognition languages
Supports PDF, DCX, DJV, DJVU, JPG, JPEG, J2C, J2K, JP2, PNG, TIF, TIFF, BMP, PCX	Supports PDF, DCX, DJV, DJVU, JPG, JPEG, J2C, J2K, JP2, PNG, TIF, TIFF, BMP, PCX
Generates four types of PDF files, PDF-iHQC (level I), four types of XPS, XPS-iHQC (level I), DOCX, ODT, XLS, WordML, SpreadsheetML, RTF, HTM, XML, TXT, TIFF, etc.	Generates four types of PDF files, PDF-iHQC (level I-III), PDF/A, four types of XPS, XPS-iHQC (level I), DOCX, ODT, XLS, WordML, SpreadsheetML, RTF, HTM, XML, TXT, TIFF, etc.
	Large volume recognition
	Automated processing
	Document indexing
	Business card recognition

Readiris Pro 12 Asian	Readiris Corporate 12 Asian
Basic features	Basic features
128 recognition languages	128 recognition languages
Supports PDF, DCX, DJV, DJVU, JPG, JPEG, J2C, J2K, JP2, PNG, TIF, TIFF, BMP, PCX.	Supports PDF, DCX, DJV, DJVU, JPG, JPEG, J2C, J2K, JP2, PNG, TIF, TIFF, BMP, PCX.
Generates four types of PDF files, PDF-iHQC (level I), four types of XPS, XPS-iHQC (level I), DOCX, ODT, XLS, WordML, SpreadsheetML, RTF, HTM, XML, TXT, TIFF, etc.	Generates four types of PDF files, PDF-iHQC (level I-III), PDF/A, four types of XPS, XPS-iHQC (level I), DOCX, ODT, XLS, WordML, SpreadsheetML, RTF, HTM, XML, TXT, TIFF, etc.
Traditional and Simplified Chinese recognition	Traditional and Simplified Chinese recognition
Japanese recognition	Japanese recognition
Korean recognition	Korean recognition
	Large volume recognition
	Automated processing
	Document indexing
	Business card recognition

Readiris Pro 12 Middle-East*	Readiris Corporate 12 Middle-East*
Basic features	Basic features
128 recognition languages	128 recognition languages
Supports PDF, DCX, DJV, DJVU, JPG, JPEG, J2C, J2K, JP2, PNG, TIF, TIFF,	Supports PDF, DCX, DJV, DJVU, JPG, JPEG, J2C, J2K, JP2, PNG, TIF, TIFF,

<p>BMP, PCX.</p> <p>Generates four types of PDF files, PDF-iHQC (level I), four types of XPS, XPS-iHQC (level I), DOCX, ODT, XLS, WordML, SpreadsheetML, RTF, HTM, XML, TXT, TIFF, etc.</p> <p>Arabic and Farsi recognition</p> <p>Hebrew recognition</p>	<p>BMP, PCX.</p> <p>Generates four types of PDF files, PDF-iHQC (level I-III), PDF/A, four types of XPS, XPS-iHQC (level I), DOCX, ODT, XLS, WordML, SpreadsheetML, RTF, HTM, XML, TXT, TIFF, etc.</p> <p>Arabic and Farsi recognition</p> <p>Hebrew recognition</p> <p>Large volume recognition</p> <p>Automated processing</p> <p>Document indexing</p> <p>Business card recognition</p>
---	--

(*) No Mac version available

CHAPTER 2

INSTALLING READIRIS

SYSTEM REQUIREMENTS

This is the **minimal system configuration** required to use Readiris:

- a 486-based Intel PC or compatible. A Pentium-based PC is recommended.
- 256 MB RAM.
- 120 MB free disk space.

(105 MB of disk space suffices when you do not install the sample files)

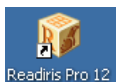
- the Windows Vista, Windows XP or Windows 2000 operating system.

Note that some scanner drivers may not work under the latest version(s) of Windows. See the documentation supplied with your scanner to find out which platforms are supported.

SOFTWARE INSTALLATION

To install the software:

- Log on to Windows as administrator or make sure you have the necessary administration rights.
- Connect your scanner to your PC and install the corresponding software.
Test your scanner. If you experience any problem contact your scanner manufacturer.
- Insert the Readiris CD-ROM in the CD-ROM drive and follow the on-screen instructions to install the software.
- Click **Readiris** to start the installation.
- Select the installation language and click **OK**.
- Accept the terms of the license agreement.
- A **complete** and a **custom** installation are offered. Select the required options and click **Next** each time you are ready to go to the next screen.
All lexicons and sample images will be installed by default, as well as an electronic user guide and online help.
- Click **Finish** to complete the installation.
The submenu **I.R.I.S. Applications - Readiris** on the Windows **Programs** menu is created automatically by the installation program.
The installation program also creates a shortcut to the Readiris application on the Windows desktop.



- Repeat the installation process to install any additional software from the CD-ROM.

UNINSTALLING THE SOFTWARE

There is only one correct way to uninstall Readiris: by using the Windows (un)install wizard. You are strongly recommended *not* to uninstall Readiris or any of its software modules by manually erasing the program files.

To uninstall Readiris:

- Close the application.
- On the **Start** menu, click **Control Panel**.
- Under the **Programs** icon, click **Uninstall a program**.
- Select **Readiris** in the list and click the **Uninstall** button.
- Follow the on-screen instructions.

SOFTWARE REGISTRATION

Remember to register your Readiris license.

To register:

Use the **Registration wizard** on the **Register** menu. Follow the on-screen instructions.

PRODUCT SUPPORT

To get technical support on Readiris, please visit our website at **<http://www.irislink.com/support>**.

CHAPTER 3

GETTING STARTED

RUNNING READIRIS

To run Readiris:

- Start **Readiris** from the Windows **Start** menu or double-click the shortcut on your desktop.



- Click anywhere in the startup screen to launch Readiris.

The OCR Wizard automatically opens.

USING THE OCR WIZARD

The OCR Wizard allows you to define all the settings needed to operate Readiris efficiently.

When you start Readiris, click anywhere in the startup screen to start the OCR Wizard.

Step 1

Select the **image source**.

You can capture images using your **scanner** or open **image files**.

Select the **rotation** and **deskewing** options you want to use.

For more information, see the section **Selecting the options**.

To familiarize yourself with Readiris, use the sample images provided with the software. They can be found on the Readiris CD-ROM and in the subfolder **Samples** of the Readiris installation folder.

Click **Next** to go to the next step.

Step 2

In case you selected a scanner, click the **Change** button to select the scanner settings.

For more information on the scanner settings, see the section **Scanning paper documents**.

Click **OK** to save the settings.

Click **Next** to go to the next step.

Step 3

Click the **Change** button to change the document language. The document language is set to American English by default.

Select the required language or language combination and secondary languages in the list and click **OK**.

Use the slider to set the required **Speed-Accuracy** settings.

For more information, see the section **Selecting the document language**.

Click **OK** to save the settings.

Click **Next** to go to the next step.

Step 4

Click the **Change** button to change the output format or target application. The default target application is Microsoft Word.

Select the required output format or application in the **Send to** or **External file** list.

Click the various tabs and select the options of your choice. Options that are unavailable for the chosen format/application appear dimmed.

For more information, see the chapter **Formatting and saving documents**.

Click **OK** to save the settings.

Click **Next** to go to the next step.

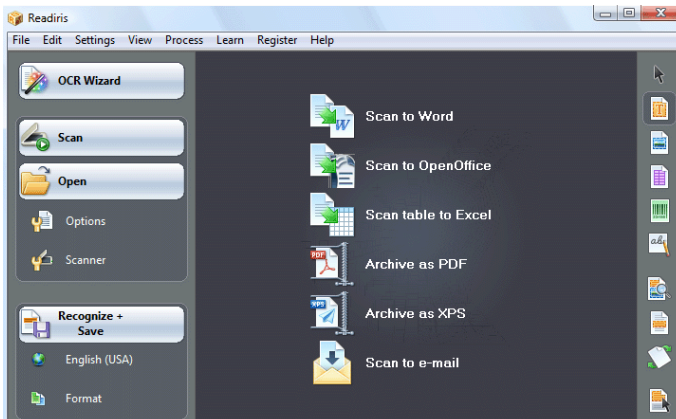
Step 5

Click **GO** to open/scan and recognize the document.

USER INTERFACE

To explore the Readiris interface, click anywhere in the **Readiris** startup screen and click **Cancel** when the OCR Wizard launches.

The empty Readiris interface will be displayed.



The Readiris interface is composed of:

- the **SmartTasks** (in the middle)

The SmartTasks are predefined commands that allow you to use the most frequent Readiris functions at the touch of a button.

Click the SmartTask you want to use to scan, recognize and send your documents to the target application or output format of your choice.

The SmartTasks apply default settings but can be configured easily by right-clicking to fit more particular needs.

- the **main toolbar** (left toolbar)

Use the main toolbar commands and options to scan and recognize documents manually.

The order in which you are advised to do so is given in the **OCR Wizard**.

- the **image toolbar** (right toolbar)

Use the image toolbar buttons to edit documents in the Readiris interface.

Point to the different buttons to display their tooltips.

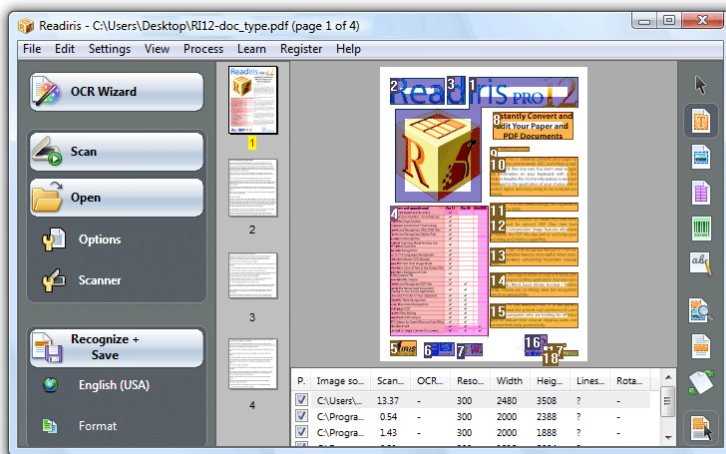
When a document has been opened or scanned in Readiris, three main zones are added to the interface:

- the **page toolbar** (right of the main toolbar)

The page toolbar displays the page thumbnails, which provide settings information if pointed to.

- the **image window** (in the middle)
- the **document panel** (at the bottom)

The document panel displays statistical information about the documents that are open in Readiris, such as the scan and OCR time, the resolution, width and height of the documents etc.



CHANGING THE USER INTERFACE LANGUAGE

The user interface of Readiris is available in a wide range of languages.

To change the user interface language:

- On the **Settings** menu, click **User Interface Language**.
- In the **Language** list, select the required language, then click **OK** to confirm.

Note: If you selected an incorrect language, click **Ctrl+U**. The **Language** dialog box will open and you will be able to select another language in the list.

CHAPTER 4

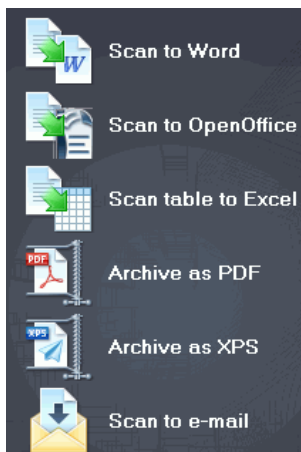
THE READIRIS SMARTTASKS

USING THE READIRIS SMARTTASKS

When starting Readiris, click anywhere in the Readiris startup screen and click **Cancel** when the OCR Wizard launches. The Readiris SmartTasks will be displayed.

The SmartTasks are predefined commands that allow you to use the most frequent Readiris functions at the touch of a button.

Simply click the SmartTasks to scan documents or image files to the target applications and output formats of your choice.



The various SmartTask buttons allow you to:

1. Scan and recognize documents and send them directly to **Word** for text processing;

Microsoft Word is the default target application. See the section **Formatting text documents** to learn more about the other available applications.

2. Scan and recognize documents and send them directly to **OpenOffice** for text processing;

OpenOffice.org Writer is the default target application. See the section **Formatting text documents** to learn more about the other available applications.

3. Scan and recognize tables and send them directly to **Excel** and other spreadsheets;

Microsoft Excel is the default target application. See the section **Formatting table-based documents** to learn more.

4. Scan and recognize documents and **archive** them as **PDF** files;

Adobe Acrobat PDF Image-Text is the default output format. See the section **Creating PDF documents** to learn more about the other available formats.

5. Scan and recognize documents and **archive** them as **XPS** files;

XPS Image-Text is the default output format. See the section **Creating XPS documents** to learn more about the other available formats.

6. Scan and recognize documents and send them directly by **e-mail**.

The documents will be sent as PDF Image-Text by default via your default e-mail application. See the section **Formatting documents** to learn more about the other available formats.

Note that the SmartTasks apply predefined settings but can be configured easily to fit more particular needs.

To configure the SmartTasks:

- Right-click the SmartTask you want to use.
- Select **Scanner** or **Image files** as image source.
 - When you select **Scanner**, Readiris will start your scanner as soon as you click the SmartTask. The scanned document(s) will be displayed in the interface, processed and saved.

Your scanner must be configured correctly in order for the SmartTasks to work.

To do so:

- Click the **Scanner** button on the main toolbar.
- Click **Scanner** model and select your scanner in the list.

If your scanner is not in the list, select **Twain other models**.

- Click **Configure** if applicable to select the Twain source.
- Then click **OK** to save the settings.

For more information on the scanner settings and on scanning paper documents, see the section **Scanning paper documents**.

- When you select **Image files** and click the SmartTask, Readiris opens the **Input** dialog box in which you can select the image files you want to process.

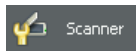
For more information on opening image files, see the section **Opening image files**.

- Click **Configure** to change the output format and its options.

Note that the available output formats and options depend on the selected SmartTask.

See the chapter **Formatting and saving documents** to learn more about the available formats and options.

- When you are done configuring the SmartTasks, use the buttons on the main toolbar to specify the **language** settings and image enhancement **options**, and if still needed the **Scanner settings**.



For more information on the above-mentioned settings, see the sections **Selecting the options**, **Scanning paper documents** and **Selecting the document language**.

- Finally, click the SmartTask to use it.

Readiris will go through the entire recognition process automatically.

CHAPTER 5

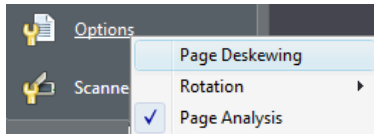
SCANNING DOCUMENTS

SELECTING THE OPTIONS

Before scanning paper documents or opening image files, you can select several image enhancement options. When enabled, these options will be applied during the opening and scanning of documents.

Operation

- Click the **Options** button on the main toolbar to select several image enhancement options.



- Click **Page Deskewing** to straighten pages scanned at an angle.
- Point to **Rotation** and determine whether you want Readiris to rotate pages **automatically** or **90° to the left**, **90° to the right** or **180°**.

Note that these two options slow down the scanning process somewhat. Only select them when necessary.

- **Page Analysis** is enabled by default.

This way, scanned or opened images will be split up in windows automatically.

You can also use the windowing tools on the image toolbar to modify the page analysis results or to window documents manually. For more information, see the chapter **Windowing documents**.

- When you are done defining all the settings (**Scanner settings, Options**), click the **Scan** or **Open** button to scan documents or open image files.

Note that the above-mentioned options are also available on the **Settings** menu.

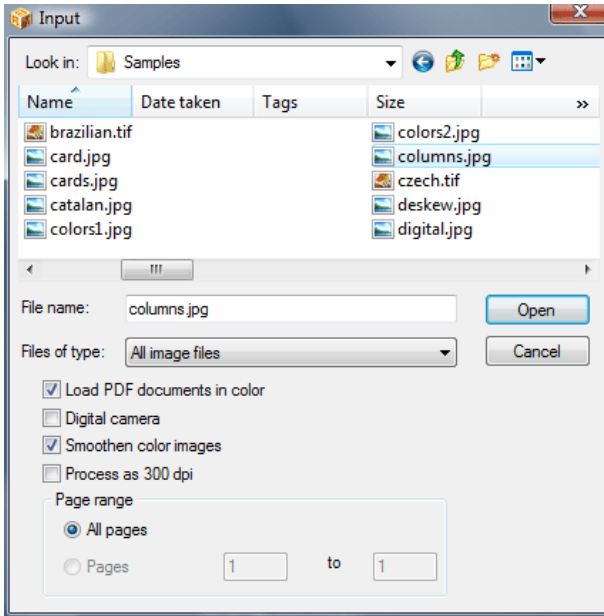
OPENING IMAGE FILES

With Readiris you can either process paper documents you scan with your scanner or process already existing images files of various formats.

To open existing image files:

- Click the **Open** button to search for image files.





Tip: you can also **drag** image files to the Readiris image window to open them.

Tip: Right-click any image file you want to open, point to **Open With** and click **IOCR application**. The Readiris software will open and display the image.

Tip: when loading multipage image files (TIFF images and DCX faxes) and PDF documents, you can define the page range (in case you only need a certain chapter of a document for instance). To do so, click **Open** on the main toolbar. In the **Page range** area, select **Pages** and enter which pages you want to load. See also **Opening and recognizing multiple image files**.

Tip: to speed up the loading process, click the **Open** button and deselect **Load PDF documents in color** when processing PDF documents.

- Readiris supports the following graphic formats:

Graphic format	File extension
Adobe Acrobat PDF	*.pdf
DCX fax	*.dcm
DjVu images	*.djv, *.djvu
JPEG images	*.jpg, *.jpeg
JPEG 2000 images	*.j2c, *.j2k, *.jp2
Portable Network Graphics	*.png
(Compressed and multipage) TIFF images uncompressed and LZW, PackBits, Group 3, Group 4 and JPEG compressed	*.tif, *.tiff
Windows bitmap	*.bmp
ZSoft Paintbrush images	*.pcx

- Select the image file of your choice and click **Open**.

Note: the options of the **Input** dialog box also apply to document scanning and are discussed in the **Scanning paper documents** section.

Note that you can specify other settings before opening (or scanning) documents. For more information, see the sections below.

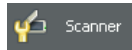
SCANNING PAPER DOCUMENTS

With Readiris you can either process paper documents you scan with your scanner or process already existing images files of various formats.

When you process paper documents, Readiris will start your scanner as soon as you click the **Scan** button and display the scanned document in the interface.

To scan documents:

- Click the **Scanner** button to set the scanner settings.



Note that several of the options in the **Scanner** dialog box are also available in the **Open** dialog box.



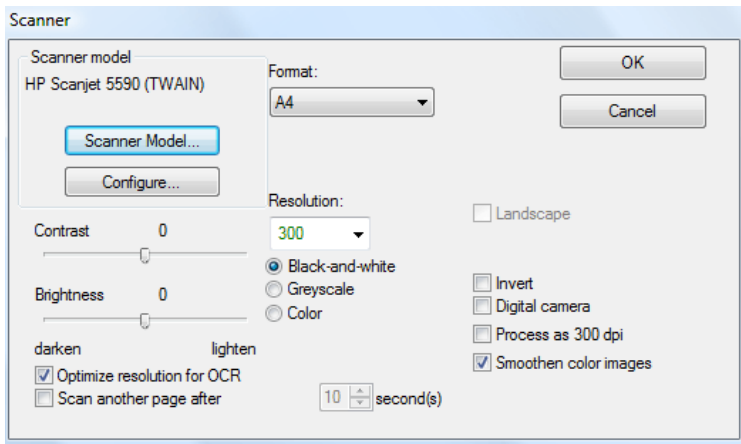
- Select the correct scanner model.

If your scanner is not in the list, select **Twain other models** and click **OK**.

- Readiris supports almost every flatbed and sheet-fed scanner, "all-in-one" device ("MFP" - "Multifunctional Peripheral"), digital camera and scanner standard currently available. Readiris is Twain compliant.

Note that the **Configure** button is only available when applicable. Click it to select the Twain source.

- Select the scanner settings.



Format and Resolution

Readiris supports a wide range of paper formats and resolutions.

Note that it is recommended to use a scan resolution of 300 dpi.

Use a resolution of 400 dpi when recognizing business cards, Asian text or very small print.

Color mode

Readiris can scan documents and open image files in **color**, **black-and-white** and **grayscale**.

Contrast - Brightness

Use the slider to determine the appropriate brightness and contrast settings, in order to obtain an optimal scan result.

Optimizing resolution for OCR

Select **Optimize resolution for OCR** to correct the resolution of images scanned with too much detail (over 600 dpi). Readiris will reduce the resolution.

Note that this option never increases the resolution of images scanned with too little detail.

Scanning multipage documents

When scanning multipage documents and using a scanner equipped with a document feeder, select the **ADF** (automatic document feeder) option. Place the pages you want to scan in the feeder and start scanning.

Or use **interval scanning** when using a flatbed scanner: select the option **Scan another page after** and indicate after how many seconds you want Readiris to scan another page.

For more information, see the section **Scanning and recognizing multipage documents**.

Important: any options that are unavailable for the selected scanner appear dimmed.

Scanning landscape images

Select the **Landscape** option when scanning landscape-oriented images.

Auto-exposure

With some scanners the option **Auto-exposure** is selected by default. This option adjusts the contrast and brightness settings automatically.

Scanning inverted images

Select the **Invert** option when scanning inverted black-and-white images.

Using a digital camera

Select **Digital camera** when you are using a camera as scan source. Readiris uses special recognition routines to process digital camera images.

Tips for using a digital camera as scan source:

- Calibrate the camera by photographing a white document.
- Always select the highest image resolution.
- Enable the macro mode of the camera to take close-ups.
- Only use optical zoom, not digital zoom.
- Hold the camera directly above the document. Avoid photographing the document at an angle.
- Produce stable images. Use a tripod if necessary.
- Disable the flash when capturing glossy paper.
- Avoid opening compressed camera images.
- Adapt the Readiris brightness and contrast settings to the environment (day light, lamp light, neon light).
- Select color or grayscale as color mode.

Processing as 300 dpi

Select **Process as 300 dpi** when you are processing images of an incorrect or unknown resolution. The images will be processed as if they had a 300 dpi resolution.

The resolution of digital camera images is nearly always unknown.

Smoothing color images

This option is selected by default as image smoothing is needed with some scanners to recognize color and grayscale images successfully.

- When you are done defining all the settings (**Scanner settings, Options**), click **Scan** to scan documents.

Note: pay attention to line skew. Line skew over 0.5° increases the risk of OCR errors.

CHAPTER 6

ADJUSTING SCANNED DOCUMENTS

When opening or scanning extremely light or extremely dark grayscale and color images, it may be necessary to adjust those images before executing the recognition, in order to obtain satisfactory OCR results.

To adjust images:

- Open or scan a color-grayscale document.

Make sure that the **scanner settings** are correct.

Important: the **scanner settings** and **adjustment settings** appear to be the same but note that both sets of settings are applied at different stages in the scan/recognition process.

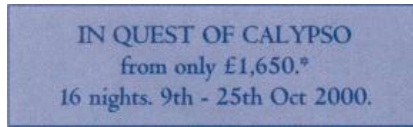
- On the **Process** menu, click **Adjust image**.

Readiris uses intelligent **binarization** routines to convert color-grayscale images into black-and-white images, which are used to perform OCR on.

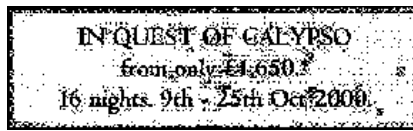
- Select **Smoothen color image** to even out the image.

This option renders grayscale and color images more homogeneous by smoothing out differences in intensity. As a result, a stronger contrast is created between the foreground (text) and background (artwork).

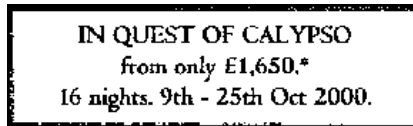
Note: sometimes smoothing is the only way to separate text from a colored background.



(Original image)



(Binarized black-and-white image)



(Smoothened image)

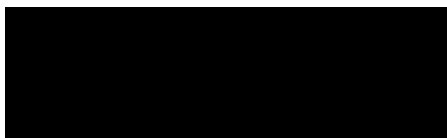
- Use the slider to increase or decrease the **Brightness**.

The **Brightness** settings determine the overall brightness of the image. Use these settings to darken or lighten the image when the text is illegible.

Example 1: lighten a dark image to eliminate the page background.



(Color image)



(Binarized image. The default binarization settings yield a black image)

**Verenigde Staten,
een antwoord te vi**

(The lightened image yields satisfactory recognition results)

Example 2: darken an image when the text is so light it doesn't show up in the binarized image.

wyjścia każdego
prawia, że nasze

(Color image)

wyjścia każdego
brawia, że nasze

(Binarized image. The default brightness settings yield fragmented characters)

wyjścia każdego
brawia, że nasze

(The darkened image yields satisfactory recognition results)

- Use the slider to increase or decrease the **Contrast**.

The **Contrast** settings determine the contrast between darker and lighter zones of an image. Use these settings to make character shapes stand out against a colored background.

A Look at International
Planning the Future

(Color image)

A Look at International
Planning the Future

(Default contrast settings yield broken characters)

A Look at International
Planning the Future

(Increased contrast settings yield satisfactory recognition results)

- Use the slider to increase or decrease the **Despeckle** options.

Despeckling removes small spots from black-and-white images.

- Click **Apply** to preview the results.
- If the results are satisfactory, click **OK**. If not, change the settings again.
- Click **Recognize + Save** to recognize the document.

CHAPTER 7

SAVING DOCUMENTS AS IMAGE FILES

Paper documents you scan do not need to be OCRed right away. They can be saved as image files.

To do so:

- Scan the document.
- On the **File** menu, click the commands **Save Full Page as Image** or **Save All Pages as Image**.
- Afterwards, open the saved image file and perform the recognition.

Saving graphics only

You can also choose to save the graphics windows without the text of the document.

To do so:

- Scan or open the document.
- On the **File** menu, click **Save Graphics**.
- All the graphics of the document will be saved in a single file.

CHAPTER 8

WINDOWING DOCUMENTS

WINDOWING DOCUMENTS AUTOMATICALLY

When scanning or opening documents, Readiris will automatically apply **Page Analysis** to split up the documents in different windows.

The **Page Analysis** option is selected by default. Click the **Options** button and disable **Page Analysis** should you want to avoid automatic page analysis.

The page analysis results can be modified manually after automatic page analysis. For more information, see the section **Windowing documents manually**.



Window types

Readiris uses five window types: **text blocks**, **graphic zones**, **tables**, **barcode zones** and **handprinted zones**.

Page analysis detects text, graphic and table zones automatically. Barcode zones and handprinted zones need to be drawn manually.

For more information, see the section **Windowing documents manually**.

Each window type has its own color code: text windows are orange, graphics are purple and table windows pink. Barcode zones are green and handprinted zones blue.

The windows are sorted top-down, left to right. Numbers indicate the sort order of the windows. The sort order and window types can be changed, however. For more information, see the section **Windowing documents manually**.

Ignore text zones on page borders

When your scanner generates black borders around the actual image, page analysis tends to find zones where there's only noise.

To avoid this, click **Do Not Detect Windows on Borders** on the **Settings** menu and scan the document again.

Ignore exterior zone

As an alternative to windowing documents automatically, the function **Ignore exterior zone** can be used. This function is useful when only one particular area on the document pages needs to be OCRed.

Select **Ignore exterior zone** by clicking the corresponding button on the image toolbar.



Draw a frame around the part of the page you want Readiris to recognize.

The part of the page you select will be analyzed automatically.

You will be prompted whether you want to exclude the same outer zone from page analysis on every page of the document.

WINDOWING DOCUMENTS MANUALLY

Besides windowing documents automatically by means of Page Analysis, Readiris allows you to window documents manually.

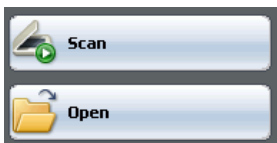
Manual windowing comes in handy when having to modify the automatic page analysis results. It is also useful to use when creating windowing templates.

For more information on windowing templates, see the section **Using windowing templates**.

Note that barcode and handprinted text zones always need to be windowed manually.

Operation

- In order to window a document manually, first click the **Options** button and deselect **Page Analysis**.
- Open or scan the document by clicking the **Scan** or **Open** button.



- Select the correct **window type** by clicking the corresponding window type button on the image toolbar.

Readiris uses five window types: **text blocks**, **graphic zones**, **tables**, **barcode zones** and **handprinted zones**.



- Draw a frame around the text blocks, graphics, tables, barcodes and handprinting zones you want to window.

For more information on recognizing barcodes and handprinting, see the sections **Recognizing barcodes** and **Recognizing handprinted text**, respectively.

- When you are done windowing the document, click the **Recognize + Save** button to execute the OCR.

Sorting windows

To change the sort order of windows, click the **Sort** button on the image toolbar and click the windows one by one in the required order.



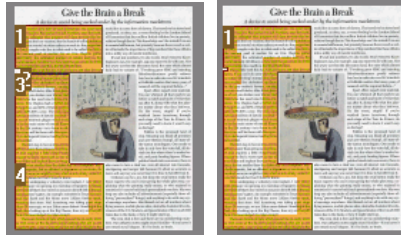
When you are done, click the **Recognize + Save** button to execute the OCR.

Windows you do not click, will be excluded from recognition.

Drawing polygons

Windowing documents manually is not limited to rectangular shapes. You can create polygonal windows by merging rectangular

ones. Whenever two windows of the same type intersect, they become a polygon automatically.



Automatic page analysis

Should the current page be too complex to window manually, click the **Analyze page** button on the image toolbar to window the page automatically.



Note that barcode zones and handprinted zones always need to be drawn manually.

Changing the window type

To change the window type of a window, right-click the window, point to **Window**, then to **Type** and then click the required window type.

You can also change the window type of several windows simultaneously:

- Click the **pointer** button on the image toolbar.



- Hold down the **Shift** key while selecting multiple windows.

- Right-click any of the selected windows, point to **Window**, then to **Type** and then click the required window type.

Modifying the window size

Click the window you want to modify.

Place the mouse pointer over a marker (on the sides and in the corners of the window).

Click the marker and drag the mouse to modify the window size.

Moving windows

Select the window you want to move.

Click inside the window and drag the mouse to modify the position of the window.

Recognizing a particular window

Right-click the window you want to recognize and select **Copy as Text**.

The results are sent to the clipboard as body text. This also works for handprinted text.

Graphic windows and barcode windows can also be copied to the clipboard.

Deleting windows

Select the window(s) you want to delete or click the command **Select All** on the **Edit** menu to select all windows.

Select the commands **Cut** or **Clear** on the **Edit** menu to cut or delete the windows.

or

Right-click the selected windows, point to **Window**, then click **Delete**.

Deleting small windows

Some documents, faxes for instance, often have "stray" dots on pages, causing Readiris to create superfluous windows that do not contain text.

To erase all small windows, click **Delete Small Windows** on the **Edit** menu.

This option erases all windows smaller than 0.5" and re-sorts the remaining zones.

USING WINDOWING TEMPLATES

When OCRing many documents with a similar page layout, it may be useful to use **windowing templates** instead of automatic page analysis. That way, the same window structure is applied to all scanned or opened documents, which speeds up the process.

Operation

- Window the first page of the document manually by using the image toolbar buttons.

For more information, see the section **Windowing documents manually**.

- On the **File** menu, click the command **Save Layout**.
- Open or scan the other pages of the document by clicking the **Open** or **Scan** button on the main toolbar.

- On the **File** menu, click the command **Load Layout**.
- Select the layout file you saved.
- To apply the layout to **all** opened or scanned pages, select **Apply Layout to All Pages** in the **Layout file** dialog box.
- Click **Open** to load the layout file.

Note that when you add a document to Readiris, the layout file must be loaded again as page analysis is enabled by default.

Ignore exterior zone

As an alternative to windowing templates, you can use the option **Ignore exterior zone**. That way, you can define one particular area on the page that needs to be OCRed. Any data outside the OCR area will be excluded from recognition.

Operation

- Select **Ignore exterior zone** by clicking the corresponding button on the image toolbar.



- Draw a frame around the part of the page you want Readiris to recognize.

The part of the page you select will be analyzed automatically.

You will be prompted whether you want to ignore the same exterior zone for all pages of the document.

To cancel this function, re-execute **Page Analysis** by clicking the **Analyze page** button on the image toolbar.



- Click **Recognize + Save** to execute the OCR.

CHAPTER 9

RECOGNIZING DOCUMENTS

INTRODUCTION

To recognize documents, Readiris applies linguistics during the recognition phase. As a result, Readiris recognizes text, tables, graphics, barcodes and handprinted text in all kinds of documents. Readiris even copes with complex columnized documents, low-quality documents, faxes, dot matrix printouts, badly scanned and copied documents containing too light or dark font shapes, etc.

Readiris supports 128 languages: all American and European languages are supported, including the Central-European, Baltic and Cyrillic languages as well as Greek and Turkish. Optionally, Readiris can read Arabic, Farsi and Hebrew documents and four Asian languages - Japanese, Simplified and Traditional Chinese and Korean. Readiris even copes with mixed alphabets: the software detects “Western” words that occur in Greek, Cyrillic, Arabic, Hebrew and Asian documents - many untranscribable proper names, brand names etc. are written using the Western symbols.

Readiris is based on the most advanced recognition technologies. Font-independent text recognition is complemented by self-learning techniques. The system is able to learn new characters and words through contextual and linguistic analysis. This means that the OCR accuracy of the recognition system will improve as it goes along.

Besides that, Readiris has an (optional) user verification function. When activated, the user verification function (**Interactive learning**) not only flags the characters the recognition system isn't

sure of but also allows to increase the system's accuracy. All solutions you confirm are memorized temporarily during recognition, increasing the system speed and confidence and rendering the system more intelligent as you go along. This powerful learning tool also allows you to train Readiris on special characters such as mathematical symbols and dingbats and to handle distorted fonts.

The interactive learning results can also be stored permanently in **font dictionaries** for future use.

SELECTING THE DOCUMENT LANGUAGE

Readiris offers OCR in 128 languages. Readiris supports all American and European languages including the Central-European, Cyrillic and Baltic languages, as well as Greek and Turkish.

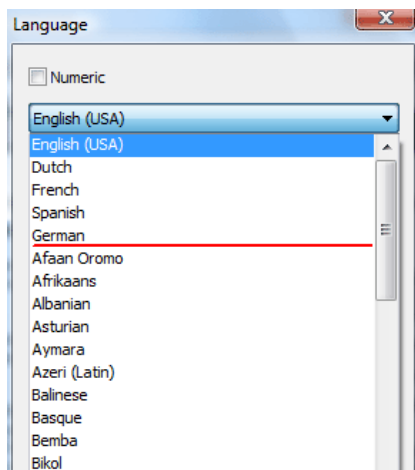
Readiris Pro Asian and Readiris Corporate Asian additionally recognize documents in Japanese, Simplified Chinese, Traditional Chinese and Korean.

Readiris Pro Middle-East and Readiris Corporate Middle-East additionally recognize documents in Arabic, Farsi and Hebrew.

In order for Readiris to recognize a document, the document language must be specified.

To do so:

Click the **globe** button on the main toolbar and select the language of your choice in the drop-down list.



The 5 most recently selected languages are moved to the top of the language list.

Important: select the document language *before* executing page analysis when you are dealing with Asian, Hebrew and Arabic documents. Specific page analysis routines are used for these documents.

The recognition can also be limited to a **numeric** character set to optimally recognize tables and figures. Readiris then only recognizes the numerals 0-9 and the following series of symbols:

+	*	/	%	,
plus sign	asterisk	slash	percentage sign	comma
.	()	-	=
period	opening parenthesis	closing parenthesis	hyphen	equation sign
\$	£	€	¥	
dollar sign	pound sign	euro sign	yen sign	

To activate numeric mode, select **Numeric** in the **Language** dialog box.

Recognizing Western words in Cyrillic and Slavic languages and in Greek

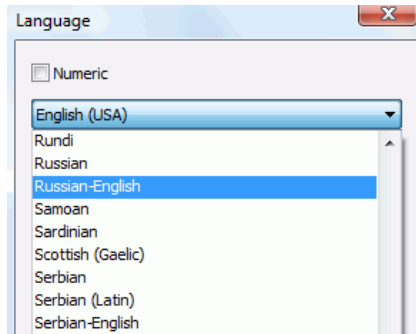
In certain Cyrillic and Slavic languages, as well as in Greek, Western proper nouns often remain written in the Latin alphabet.

Readiris is able to recognize those Western words in Russian, Byelorussian, Ukrainian, Serbian, Macedonian, Bulgarian and Greek.

You must simply indicate the language pair (Other language – English) that is used in the text. Readiris will use the corresponding character set and switch languages whenever necessary.

To enable the reading of Western words in the above-mentioned languages:

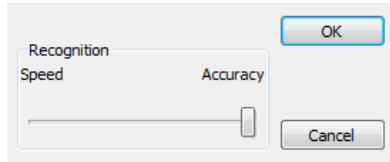
- Click the **globe** button on the main toolbar.
- Select the required language pair in the language drop-down list.



Note: when processing Asian or Hebrew documents, mixed characters sets are used automatically.

Speed-Accuracy

Select the right trade-off between **OCR speed** and **OCR accuracy**.



This trade-off is available for the Latin, Cyrillic and Greek alphabets.

Tip: favor accuracy over speed when the image quality is rather poor.

DEFINING THE DOCUMENT CHARACTERISTICS

Next to the document language, other document characteristics such as the **Font type** and **Character pitch** play an important role in the recognition process.

Font type

Readiris distinguishes between "regular" and dot matrix printed documents. Dot matrix symbols (of the type 9 pin) are made up of isolated, separate dots.

Special segmentation and recognition techniques are required to recognize dot matrix documents and need to be activated.

Far out in the uncharted back

To select the font type:

- On the **Settings** menu, point to **Font type**.

- The font type is set to **Automatic** by default.

That way, Readiris recognizes "25 pin" or "NLQ" (Near Letter Quality) dot matrix, or other "normal" printing.

- To recognize only dot matrix printed documents, click **Dot matrix**.

Readiris will recognize so-called "draft" or "9 pin" dot matrix printed documents.

Character pitch

The character pitch is the number of characters per inch in a typeface. The character pitch can either be **fixed**, in which case all characters have the same width, or **proportional**, in which case the characters have a different width.

To select the character pitch:

- On the **Settings** menu, point to **Character Pitch**.
- The character pitch is set to **Automatic** by default.
- Click **Fixed** if all characters of the typeface have the same width. This is often the case in old typewriter documents.
- Click **Proportional** if the characters of the typeface have a different width. Virtually all fonts in newspapers, magazines and books are proportional.



Important: these document characteristics do not apply to Asian, Hebrew or Arabic documents.

USING INTERACTIVE LEARNING

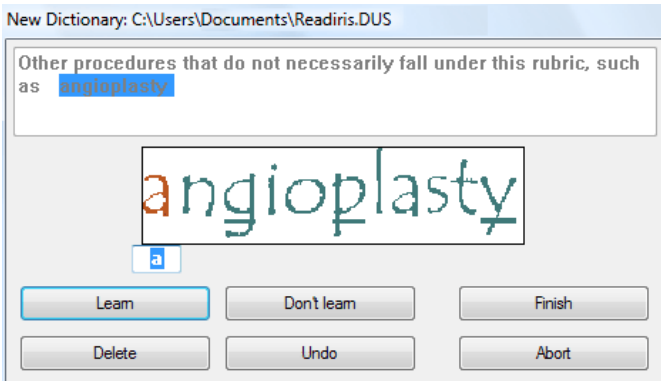
Readiris offers an interactive learning function. By means of **Interactive learning** you can train the recognition system on fonts and character shapes, and correct the OCR results if necessary. During interactive learning, any characters the recognition system isn't sure of are displayed in a preview window, in combination with their parent word and the proposed solution.

Interactive learning can substantially enhance the accuracy of the recognition system and is particularly useful when recognizing distorted, defaced forms. Interactive learning can also be used to train Readiris on special symbols it is unable to recognize initially, such as mathematical and scientific symbols and dingbats.

To enable interactive learning:

- On the **Learn** menu, click **Interactive Learning**.
- Define the necessary settings and click the **Recognize + Save** button to recognize the document.
- At the end of the recognition, Readiris enters the interactive learning phase.

The characters the recognition system isn't sure of are displayed.



If the results are correct:

- Click the **Learn** button to save the result as sure.

The learning results are temporarily stored in the computer memory, for the duration of the recognition. Readiris will no longer display the learned characters when OCRing the rest of the document.

When a new document is OCRed, the learning results are erased.

To save learning results permanently, use a font dictionary. For more information, see the section **Using font dictionaries**.

- Click **Finish** to save all solutions the software offers.

If the results are incorrect:

- Type in the correct characters and click the **Learn** button.

or

- Click **Don't learn** to save the result as unsure.

Use this command for damaged characters which could be confused with other characters if learned. E.g. the number 1 and the letter I, which have an identical form in many fonts.

- Click **Delete** to delete characters from the output.

Use this button to prevent document noise from appearing in the output file.

- Click **Undo** to correct mistakes.

Readiris keeps track of the last 32 operations.

- Click **Abort** to abort interactive learning.

All learning results will be deleted. Next time you click **Recognize** + **Save**, interactive learning will start again.

USING FONT DICTIONARIES

When scanning many documents of the same type, font quality and printing quality, you may not want to repeat the learning process every time. Therefore, it is useful to use font dictionaries. Font dictionaries contain font information learned during **interactive learning** and can substantially increase the recognition results.

Note that font dictionaries are limited to 500 shapes. You are recommended to create separate dictionaries for specific applications.

To create a new font dictionary:

- On the **Learn** menu click the command **New Font Dictionary**.

Readiris will open the dictionary **Readiris.dus** by default. Change the file name and click **Save** to save it as a new dictionary.

- Click **Interactive Learning** on the **Learn** menu to activate it.
- Click **Recognize** + **Save** to recognize the document.
- Readiris enters the interactive learning phase. Use the buttons of the dialog box to save characters in the font dictionary.

To use an existing font dictionary:

- On the **Learn** menu click **Font Dictionary**.
- Select the dictionary you want to use and click **Open**.
- On the **Learn** menu click either **Append Font Dictionary** or **Read Font Dictionary**.

When selecting **Append Font Dictionary**, make sure to enable **Interactive Learning**. Readiris will recognize the character shapes stored in the dictionary and use interactive learning, allowing you to store new information in the dictionary.

When selecting **Read Font Dictionary**, Readiris will recognize the character shapes stored in the dictionary but will not add new content to the dictionary, even if **Interactive Learning** is enabled. Note that it is still useful to use **Interactive Learning** to check, and if necessary, correct the recognition results which are not saved in the font dictionary.

Caution: do not click **Font Dictionary** on the **Learn** menu and open an existing dictionary while the dictionary mode **New Dictionary** is enabled. Otherwise, the contents of the existing font dictionary will be erased.

- Click **Recognize + Save** to recognize the document.

CHAPTER 10

FORMATTING AND SAVING DOCUMENTS

FORMATTING DOCUMENTS

The documents you OCR in Readiris can be saved in various output formats. Readiris saves OCR results as Adobe Acrobat PDF files, Microsoft XPS files, Word, WordML, RTF and OpenDocument text files, HTML and XML files, SpreadsheetML worksheets, and Ansi and Unicode text files.

Besides that, Readiris can export results directly to such target applications as Microsoft Word and Excel, Adobe Reader, Microsoft XPS Viewer, the major web browsers and e-mail software etc.

Note: RTF files created by Readiris can also be opened in the web-based office applications AjaxWrite, ThinkFree, Zoho Writer and Google Writely (which opens both RTF and HTML documents). When using one of these applications make sure to select the layout option **Retain Word and Paragraph Formatting**.

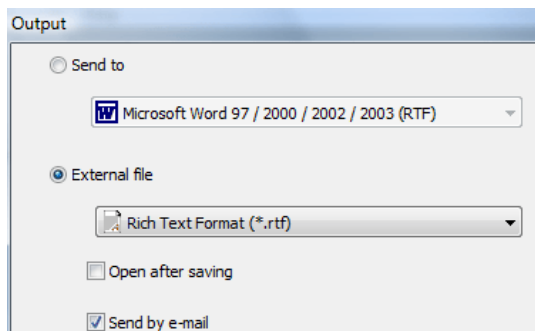
Operation

- Click the **Format** button on the main toolbar to select the output mode.

Readiris either:

- sends documents to an application, which will open automatically, or;
- saves documents as an external file.

The option **Send by e-mail** creates a new e-mail message and inserts the recognized document as e-mail attachment.



- Click the different tabs to select the settings you want to apply.

Settings that are unavailable for the selected output format appear dimmed.

The most commonly used output file formats as well as their options are discussed in the sections below.

- Click **Recognize + Save** to execute the recognition and save the documents.

The OCR results can be exported several times without repeating the recognition. Click the **Format** button again and change the text format and formatting options. Then click **Recognize + Save** again.

For searching and sorting reasons, Readiris allows you to define **document properties** of PDF, XPS, Word, RTF, WordML, SpreadsheetML and HTML output. To define the document properties of a document, click **Document Properties** on the **File** menu. Note that the document properties options are also accessible in the **Output File** dialog box, which opens when you click **Recognize + Save**.

Note that when saving a multipage document as **external file**, you can create a separate output file for each page in Readiris or save all pages that belong to the same document to a single output file. Simply click the corresponding options in the **Output File** dialog box: **Create one file per page** and **Create one file per document**, respectively.

Note, however, that the options **Create one file per page** and **Create one file per document** are only available when saving documents as an external file, not when opening documents in a target application.

FORMATTING TEXT DOCUMENTS

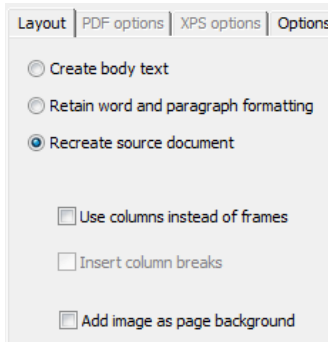
With Readiris you can generate several types of text-based output formats. Readiris offers a.o. Word, WordML, RTF, txt and OpenDocument Text output.

To generate text-based output files:

- Click the **Format** button on the main toolbar and select the output format of your choice in the **Send to** or **External file** drop-down list.
- Depending on the text format you selected, several formatting options are available.

Any options that are unavailable for the selected text format appear dimmed.

Layout options



- The option **Create body text** avoids text formatting by Readiris.

Readiris generates a continuous, running text.

- The option **Retain word and paragraph formatting** takes an intermediate position between body text and autoformatting.

The font type, size and type style are maintained across the recognition.

The tabs and the alignment of each block are recreated.

The text blocks and columns aren't recreated; the paragraphs just follow each other.

The tables are recaptured correctly.

- The option **Recreate source document** recreates a facsimile copy of the original document.

Readiris generates a true copy of the source document, no longer a scanned image.

Readiris also recreates any hyperlinks to e-mail addresses and web sites.

- The option **Use columns instead of frames** creates columnized documents.

Columnized texts are easier to edit than documents containing multiple frames: the text flows naturally from one column to the next.

Note: when the system is unable to detect columns in the source document, this formatting mode uses frames as a fallback position.

- The option **Insert column breaks** inserts a hard column break at the end of each column.

Any text you edit, add or remove, remains inside its column; no text ever flows automatically across a column break.

Tip: disable this option when you have columnized body text. You'll ensure the natural flow of the text from one column to the next.

- The option **Add image as page background** places the scanned image as page background beneath the recognized text.

This option increases the file size of the output files substantially, however.

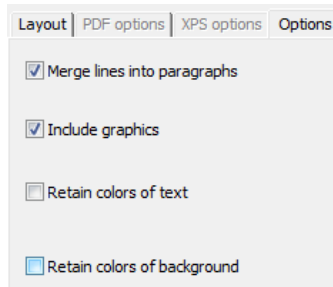
The format **PDF Text-Image** provides the same result for PDF files.

The option **Retain colors of background** on the **Options** tab provides a less drastic, more compact alternative.

General options

Click the **Options** tab to select the general options:

T



- The option **Merge lines into paragraphs** enables automatic paragraph detection.

Readiris wordwraps the recognized text until a new paragraph starts, and "reglues" hyphenated words at the end of a line.

- The option **Include graphics** includes the graphics in autoformatted files.

This is essential to create a true copy of a document.

Use the graphic options on the **Graphics** tab to determine the color mode and resolution of the graphics stored inside the output files.

- The option **Retain colors of text** maintains the original colors of the text across the recognition.
- The option **Retain colors of background** maintains the spot colors of the page background across the recognition.

A uniform background color is created per paragraph in the output file.

Paper sizes

Depending on the format you selected, you can select preferred paper sizes:

- Click the **Paper size** tab and use the arrow buttons to apply and exclude paper sizes.
- Readiris will go through the active paper sizes in the indicated order and will use the first paper size that is sufficiently large to hold the scanned document.

FORMATTING TABLE-BASED DOCUMENTS

With Readiris you can output tabular data to spreadsheets, word processors and web browsers: tables are reconstructed cell by cell in worksheets and inserted as table objects in word processor files.

Readiris recognizes both gridded and non-gridded tables.

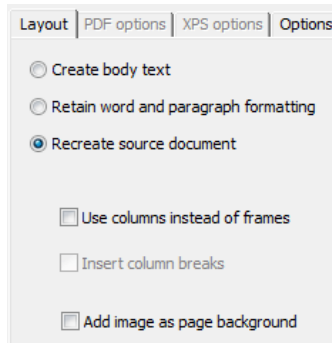
Performance test optical media				
CD-ROM	Average	123.985	69	31
Digital Versatile Disk	time (n	987.745	129	2
CD-ROM 32x speed	60			
DVD	58	19	149	91
Tested on 333 MHz Pentium II		287.410	49	52

(gridded)

(non-gridded)

To generate table-based documents:

- Click the **Format** button on the main toolbar and select the output format of your choice in the **Send to** or **External file** drop-down list.
- Select the layout options of your choice.

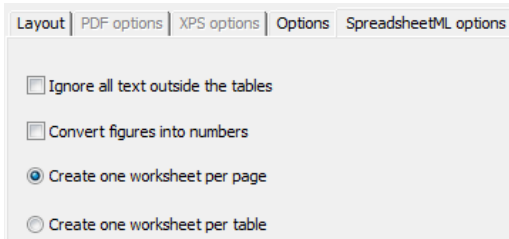


For more information on formatting options, see the section **Formatting text documents**.

SpreadsheetML options

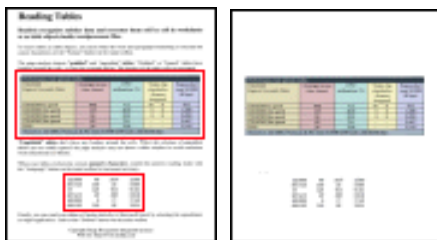
When selecting **Microsoft Excel 2002/2003** as target application, specific **SpreadsheetML** options are available. Click the tab **SpreadsheetML options** to display them:

Note that the layout option **Recreate source document** becomes unavailable when this format is selected.



- The option **Ignore all text outside the tables** saves the tables and ignores all other recognition results.

All data inside the tables is recaptured; any data outside the tables is not.



- The option **Convert figures into numbers** encodes recognized figures as numbers.

As a result, you can execute arithmetical operations on those cells. The text cells (in any table) remain text.

Note that only figures inside tables are encoded as numbers.

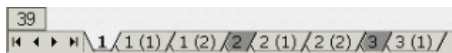
Excel exclusively executes mathematical operations on data that is encoded as numbers.

- The option **Create one worksheet per page** sees to it that one worksheet is created per scanned page.

If a page contains tables and text, all will be placed on the same worksheet.

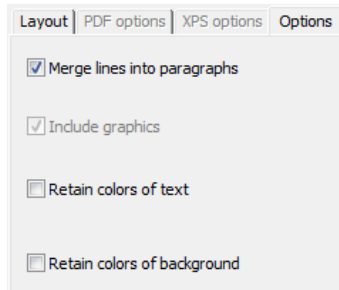
- The option **Create one worksheet per table** places each table in a separate worksheet and includes the recognized text (outside the tables) in another worksheet.

If the recognized document contains **several pages**, you'll see that structure repeated per page.



General options

Click the **Options** tab to select the general options:



- The option **Merge lines into paragraphs** enables automatic paragraph detection.

Readiris wordwraps the recognized text until a new paragraph starts, and "reglues" hyphenated words at the end of a line.

- The option **Retain colors of background** recreates the background color of each cell.

	A	B	C
1	Performance optical media		
2	CD-ROM	Average access	CPU
3	Digital Versatile Disk	time (msec)	utilization (%)
4	CD-ROM 24x speed	80	58.2
5	CD-ROM 32x speed	60	72.1
6	DVD	58	78.9
7	Tested on 333 MHz Pentium II with 64 MB RAM and 4 GB HD		

Paper sizes

Depending on the format you selected, you can indicate preferred paper sizes:

- Click the **Paper size** tab and use the arrow buttons to apply and exclude paper sizes.
- Readiris will go through the active paper sizes in the indicated order and will use the first paper size that is sufficiently large to hold the scanned document.

CREATING PDF DOCUMENTS

Readiris generates four types of PDF output: **Text**, **Text-Image**, **Image-Text** and **Image**.

To generate PDF output:

Click the **Format** button on the main toolbar and select the PDF type of your choice in the **Send to** or **External file** drop-down list:

PDF Image

When you select PDF **Image**, Readiris generates image-only PDF documents, it does not execute OCR.

PDF Image-Text

When you select PDF **Image-Text**, Readiris recognizes text and creates searchable PDF files that contain the page image and the recognized text.

The page image is placed on top of the text.

With this format you can always see the original document (as it was scanned) while you are able to search for and copy-paste the OCRed text, which is hidden beneath the image. As a result, this format is useful for **archiving purposes**.

PDF Text

When you select PDF **Text**, Readiris recognizes text and creates searchable PDF files.

The page image is not contained in these single-layered PDF files.

PDF Text-Image

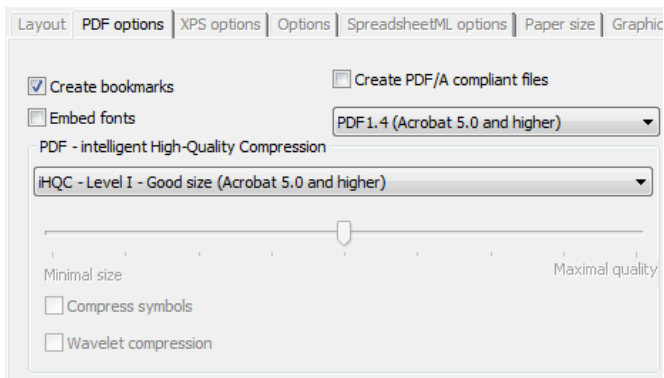
When you select PDF **Text-Image**, Readiris recognizes text and creates searchable PDF documents that contain the page image and the recognized text.

The page image is contained beneath the text.

SELECTING THE PDF OPTIONS

To select the PDF options:

- Click the **Format** button on the main toolbar and select the PDF type of your choice in the **Send to** or **External file** drop-down list.
- Depending on the **PDF type** you select, several options are available. Click the **PDF options** tab to access them:



Create bookmarks

The option **Create bookmarks** creates bookmarks for each text block, graphic and table in Adobe Acrobat PDF files.

Embed fonts

Select the option **Embed fonts** to embed fonts in Adobe Acrobat PDF files.

Embedding fonts prevents font substitution and ensures that readers, regardless of their computer configuration, see the text in its original fonts.

Embedding fonts increases the file size of recognized documents somewhat.

iHQC COMPRESSING PDF DOCUMENTS

Besides four types of "regular" PDF output, Readiris offers iHQC compressed PDF output: PDF documents of the types **Image-Text** and **Image** can be hyper-compressed by means of iHQC without loss of image quality.

iHQC stands for **intelligent High-Quality Compression**, I.R.I.S.' proprietary, efficient compression technology. iHQC is to images what MP3 is to music and what DivX is to movies.

To generate iHQC compressed PDF output:

- Click the **Format** button on the main toolbar and choose between the two output modes.
- In the **Send to** or **External file** list, select the PDF type of your choice: **PDF Image-Text** or **PDF Image**.

- On the **PDF Options** tab, select the required compression level.

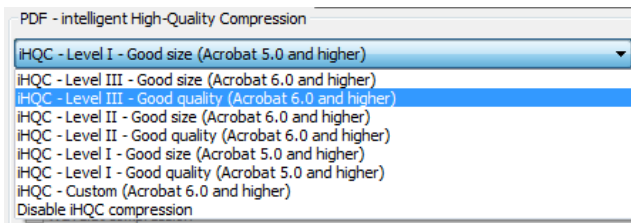
Readiris Pro supports **Level I - Good size** and **Level I - Good quality** compression.

Readiris Corporate also supports both **Level II and III Good size** and **Good quality** compression as well as **Custom** compression.

In Level II compression the option **Compress symbols** is enabled automatically to compress text compactly.

In Level III compression also the option **Wavelet compression** is enabled automatically to compress graphics compactly.

When you select **Custom** compression, you can enable or disable these options independently of one another. You can also use the slider to define the **Good size - Good quality** ratio.



- Click **Recognize + Save** to recognize the document and save it as an iHQC compressed PDF file.

REPURPOSING PDF DOCUMENTS

Next to generating PDF documents, Readiris can also repurpose PDF files: Readiris converts **image** PDFs into **text** PDFs or any other supported text format and **unlocks** read-only PDF content.

Warning: Readiris does not open user password-protected PDF documents.

Operation

- Click the **Open** button on the main toolbar and select the PDF file you want Readiris to repurpose.



In the **Page range** area of the dialog box, select **Pages** and indicate which pages you want Readiris to open.

- Click the **Open** button in the dialog box to open the PDF file of your choice.
- Click the **Format** button on the main toolbar and select the PDF type of your choice.

For more information on the PDF types, see the section **Creating PDF documents**.

- Click the **Recognize + Save** button to repurpose the document.

CREATING XPS DOCUMENTS

Readiris generates four types of XPS files: **Text**, **Text-Image**, **Image-Text** and **Image**.

XPS stands for XML Paper Specification and is a fixed-layout format developed by Microsoft.

To generate XPS output:

Click the **Format** button on the main toolbar and select the XPS type of your choice in the **Send to** or **External file** drop-down list:

XPS Image

When you select XPS **Image**, Readiris generates image-only XPS documents, it does not execute OCR.

XPS Image-Text

When you select XPS **Image-Text**, Readiris recognizes text and creates searchable XPS files that contain the page image and the recognized text.

The page image is placed on top of the text.

With this format you can always see the original document (as it was scanned) while you are able to search for and copy-paste the OCRred text, which is hidden beneath the image. As a result this format is useful for **archiving purposes**.

XPS Text

When you select XPS **Text**, Readiris recognizes text and creates searchable XPS files.

The page image is not contained in these single-layered XPS files.

XPS Text-Image

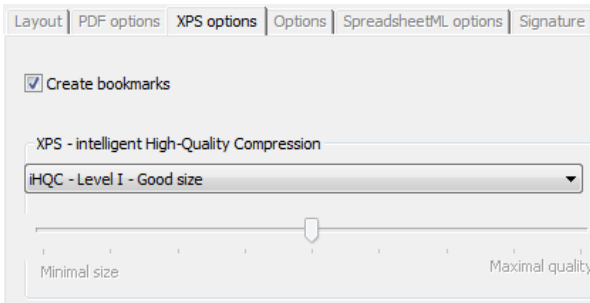
When you select **XPS Text-Image**, Readiris recognizes text and creates searchable XPS documents that contain the page image and the recognized text.

The page image is contained beneath the text.

SELECTING THE XPS OPTIONS

To select the XPS options:

- Click the **Format** button on the main toolbar and select the XPS type of your choice in the **Send to** or **External file** drop-down list.
- Depending on the **XPS type** you select, several options are available. Click the **XPS options** tab to access them:



Create bookmarks

The option **Create bookmarks** creates bookmarks for each text block, graphic and table in Microsoft XPS files.

iHQC COMPRESSING XPS DOCUMENTS

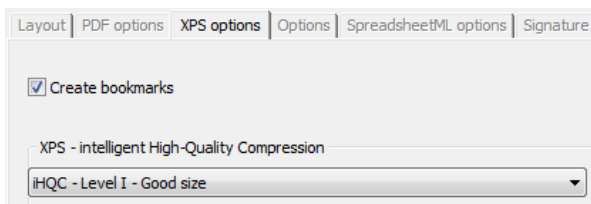
Besides four types of "regular" XPS output, Readiris offers iHQC compressed XPS output. XPS documents of the types **Image-Text** and **Image** can be hyper-compressed by means of iHQC.

iHQC stands for **intelligent High-Quality Compression**, I.R.I.S.' proprietary, efficient compression technology. iHQC is to images what MP3 is to music and what DivX is to movies.

To generate iHQC compressed XPS output:

- Click the **Format** button on the main toolbar and choose between the two output modes.
- In the **Send to** or **External file** list, select the PDF type of your choice: **XPS Image-Text** or **XPS Image**.
- On the **XPS Options** tab, select the required compression level.

Readiris supports **Level I Good size** and **Level I Good quality** compression.

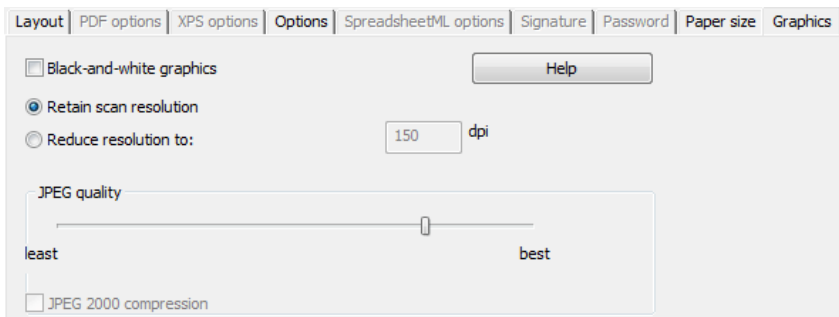


SELECTING THE GRAPHICS OPTIONS

Depending on the output format and target application you select, advanced graphics options may be available. The graphics options can be used to alter the image quality and resolution.

To access the graphics options:

- Click the **Format** button on the main toolbar and select the output format of your choice in the **Send to** or **External file** drop-down list.
- Click the **Graphics** tab to display the options.



Color mode

Readiris saves graphics in **color** by default.

Select **Black-and-white** to save graphics in black-and-white.

Resolution

Readiris **retains the scan resolution** by default.

You can also choose to **reduce the resolution**.

Tip: When saving documents as HTML files to post on a web site, reduce the resolution to 70 dpi (screen resolution).

JPEG quality

Graphics stored inside PDF, XPS, Word and RTF documents are saved in the JPEG format.

Use the slider to adjust the JPEG quality.

JPEG 2000 compression

When saving files in the PDF or XPS format, Readiris can apply JPEG 2000 compression to the color-grayscale images stored inside those files.

JPEG 2000 is the newest, more compact version of the JPEG standard.

Select the option **JPEG 2000 compression** to apply it.

CHAPTER 11

SAVING AND LOADING SETTINGS

Any settings you specify in Readiris are saved automatically for future use after you close the application. To restore the factory settings, click the command **Restore Factory Settings** on the **File** menu.

When scanning various groups of documents which all require different settings, it is useful to save separate **settings files** for each group.

Operation

- Select the settings you want to use for a certain document group.
- On the **File** menu click the command **Save Settings**.
- When scanning or opening a document of the same group at a later time, click the command **Load Settings** on the **File** menu.
- Select the correct settings file and click the **Open** button.
- Click **Recognize + Save** to recognize the document, using the correct settings.

Note: the **Info** command on the **File** menu gives an overview of the most important settings you selected.

CHAPTER 12

RECOGNIZING MULTIPAGE DOCUMENTS

OPENING AND RECOGNIZING MULTIPLE IMAGE FILES

Readiris is designed to process multiple image files at a time.

To open multiple image files:

- Click **Open** on the main toolbar.



- Select the image files you want to open:
 - Select the first image file and hold down the **Ctrl** key as you select additional images or;
 - Select a continuous range of image files by clicking the first image and holding down the **Shift** key as you select the last image.

Note: when you open a single file that consists of multiple pages, e.g. a multipage TIFF file or a PDF document, you can indicate the page range. In the **Page range** area of the dialog box, select **Pages** and indicate which pages you want to recognize.

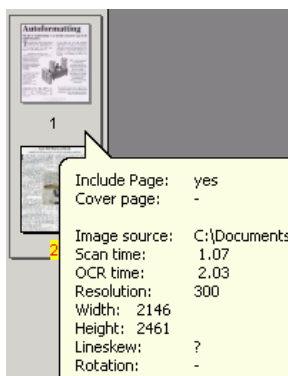
- Click the **Open** button to open the image(s).

Note that you can also drag-and-drop image files from Windows Explorer to the Readiris image window to open them.

- The page toolbar will display the opened image files.

Tip: hold the mouse cursor over the page thumbnails to display the settings information per page.

The page toolbar can be used to edit multipage documents. For more information, see the section **Editing multipage documents**.



- Determine the recognition settings and click **Recognize + Save** to execute the recognition.
- Should you want to open or scan additional images to the current document, click the **Scan** or **Open** button on the main toolbar.

You will be prompted whether you want to delete the current document or not. Click **yes** to delete the current document and start a new one or click **no** to add the additional scans to the current document.

SCANNING AND RECOGNIZING MULTIPAGE DOCUMENTS

Readiris is designed to process documents consisting of multiple pages.

Readiris Home and Readiris Pro process documents of up to 50 pages.

Readiris Corporate processes documents of an unlimited number of pages.

To scan multipage documents in Readiris, you can either use the **automatic document feeder** function when using a sheet-fed scanner or use **interval scanning** function when you are using a flatbed scanner.

Scanning multipage documents with a document feeder (sheet-fed scanner)

- Click the **Scanner** button on the main toolbar and select the **ADF** (automatic document feeder) option.
- Place the pages in your scanner's document feeder and click **Scan** to start scanning.
- Click **Recognize + Save** to recognize the documents.

Scanning multipage documents with interval scanning (flatbed scanner)

- Click the **Scanner** button on the main toolbar.
- Select **Scan another page after** and indicate the time interval using the arrow buttons.

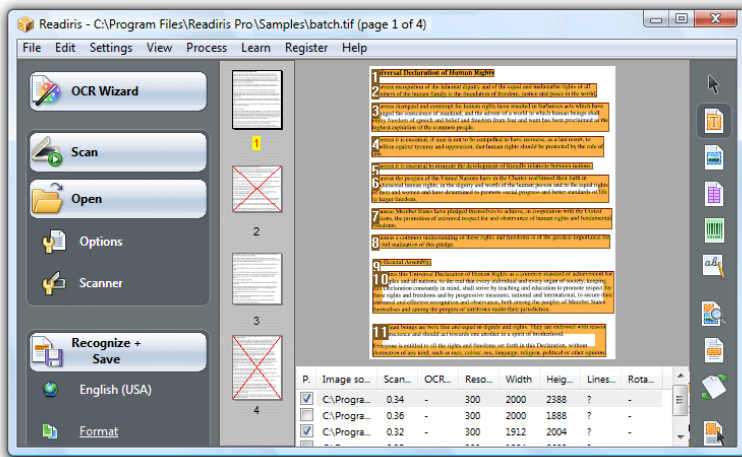
The scanner will automatically scan another page after the indicated number of seconds without you having to click the **Scan** button every time.

Click **Abort** in the interval scanning dialog box to end the automatic scanning or press **ESC** on the keyboard.

Click **Pause** in the interval scanning dialog box to freeze the scanning interval or press the space bar on the keyboard. Click **Resume** when you're ready to continue.

EDITING MULTIPAGE DOCUMENTS

When multiple documents are opened or scanned in Readiris, the page toolbar displays their thumbnails. The thumbnails in the page toolbar can be used to edit the multipage documents.



Moving a page inside a document:

- Right-click the page you want to move and click **Select Page**.

- Drag the page to the correct position.
- Or right-click a page and click **Move Page Up** or **Down**.

Deleting a page:

- Right-click the page you want to delete and click **Delete page**.
- Or select the page and hit the **Delete** button on your keyboard.

Excluding a page from recognition:

- Right-click the page you want to exclude and click **Exclude page**.
- Or clear its page number box in the document panel.

Excluded pages are stricken out in the page toolbar.



Excluded pages are ignored when you print the scanned images and when you save the scans to multipage image files.

Page	Image source	Scan time
<input type="checkbox"/> 1	C:\Docume...	5.68
<input checked="" type="checkbox"/> 2	C:\Docume...	9.45

Tip: the commands **Include All Pages** and **Exclude All Pages** on the **Edit menu** apply to all pages simultaneously.

CHAPTER 13

RECOGNIZING HANDPRINTED TEXT

Next to typed text, tables, graphics and barcodes, Readiris recognizes **handprinted** text. Handprinting consists of separated block letters.

CELL PHONE

It takes highly specialized ICR software (intelligent character recognition) to recognize handprinted characters.

To recognize handprinting:

- Click the handprinting button on the image toolbar.



- Draw a frame around the handprinted text.
- Click **Recognize + Save** on the main toolbar.

The entire document including the handprinted text will be recognized.

Note: right-click the handprinted zone and click **Copy as Text** to recognize only the handprinted zone and send it to the clipboard.

Recognized symbols

Handprinting recognition is limited to the Latin alphabet and supports numerals (0-9), uppercase letters (A-Z) and the punctuation symbols comma, period, plus sign and hyphen.

Accents, umlauts and other special characters are not supported.

Notes

- Readiris supports **handprinting**, not **handwriting**.

For more information, see the section **Handprinting rules**.

- Uppercase characters are replaced by lowercase characters after recognition, unless they occur at the beginning of a sentence.
- The document characteristics language, font type and character pitch do not apply to handprinting.
- Interactive learning does not apply either. The ICR technology is based on more than one million writing samples.

HANDPRINTING RULES

Several rules must be taken into account in order for Readiris to recognize handprinting:

- Write regular, well-spaced characters.

A B C D E F G H I
J K L M N O P Q R
S T U V W X Y Z
1 2 3 4 5 6 7 8 9 0 + , - .

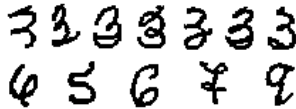
Note how the characters A, G and Q are written.

- Use a sufficiently thick ballpoint.

Black pens yield better results than blue pens.

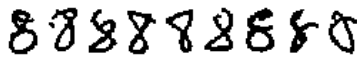
Do not use pencils.

- Don't stylize too much.



Excessively stylized characters increase the risk of OCR errors.

- Don't open loops which should be closed, don't close loops which should be open.



- Avoid broken characters.



- Avoid retracing.



Retracing reduces the image quality and clarity of handprinted symbols.

Characters that are entirely stricken out will not be recognized.

- Write ones correctly.

The number 1 can be written in the "anglicized" and "European" style.

Ones can be underlined or not.

The horizontal underlining bar does not have to touch the rest of the font form.



Tip: when less than optimal results are obtained, use the I.R.I.S. **writing form** and adapt your writing style. The blank I.R.I.S. writing form serves as a full-page template on which block letters can be filled out correctly and in the right size. The form can be found on the Readiris CD-ROM and in the Readiris installation folder.

CHAPTER 14

RECOGNIZING BARCODES

INTRODUCING BARCODE READING

Next to optical character recognition of 128 languages, Readiris also offers barcode reading.

All widespread barcode symbologies are supported: Codabar, Code 128, Code 39, Code 39 extended, Code 39 HIBC, Code 93, Datalogic 2 of 5, Discrete 2 of 5, EAN-13, EAN-8, Interleaved 2 of 5, MSI Pharmaceutical, MSI-Plessey, Kodak patch code, PDF-417, PostNet, UCC-128, UPC-A and UPC-E.



Note that laser printed and inkjet printed barcodes are required in order for Readiris to perform OCR. Matrix printed barcodes are not supported as they do not produce sufficient contrast and their resolution is mostly limited to 60 dpi.

Manual barcode reading

- Determine which barcodes you want Readiris to recognize:
 - On the **Settings** menu, click **Barcodes**.

- Select the symbologies you want Readiris to recognize.
- Determine whether you want Readiris to verify or remove the check digits.
- Click the barcode button on the image toolbar and draw a frame around the barcodes zones in the document.



- Click **Recognize + Save** on the main toolbar.

The entire document including the barcode content will be recognized.

Note: right-click a barcode zone and click **Copy as Data** to copy its content to the clipboard.

INDEX

A

accuracy vs. speed..... 55
ADF 85
adjusting scanned documents 33
Arabic documents 4, 52
Asian documents..... 7, 52
Asian edition.....4, 7, 52
automatic document feeder .. 85
automatic windowing..... 41

B

background color 65
background color of table cells
..... 70
barcodes 93
black-and-white image..... 26
brightness..... 28, 34

C

changing the user interface .. 17
character pitch..... 56

color image.....26, 33
color mode.....28
contrast28, 36

D

deskewing.....23
despeckling.....36
digital camera30
document characteristics55
document panel16
document properties62
dot matrix56

E

editing multipage documents 86
Excel output.....61

F

factory settings81
font dictionaries59
font type55

-
- G**
- graphics options 79
 - grayscale image 26
- H**
- handprinting 89, 90
 - Hebrew documents 4, 52
 - HTML output 61
- I**
- image toolbar 16
 - installation 10
 - interactive learning 57
 - interval scanning 85
 - inverted images 29
- J**
- JPEG 2000 compression 80
- L**
- language 52
 - language pairs 54
 - layout files 47
 - layout options 64
 - line skew 31
 - loading settings 81
- M**
- main toolbar 16
 - manual windowing 43
 - Middle-East edition 4, 7, 52
 - multipage documents 83, 85
- N**
- numeric 53
- O**
- OpenDocument output 61
 - options 23
 - output formats 61
- P**
- page analysis 23
 - page deskewing 23
 - pages 86
 - deleting 87
 - excluding 87
 - moving 86
 - selecting 86
 - paper sizes 66
 - PDF iHQC output 73
 - PDF options 72

PDF output..... 61, 71

product support 12

R

recreating source documents 64

registration 11

repurposing PDF documents 75

resolution 28

restoring factory settings..... 81

rotation..... 23

RTF output..... 61

S

saving settings..... 81

scanner settings..... 27

sending documents by e-mail62

settings file..... 81

SmartTasks 19

smoothing color images .. 30,
34

speed vs. accuracy..... 55

spreadsheet documents..... 67

SpreadsheetML output 61

supported image formats25

system requirements9

T

tables67

text documents.....63

U

Unicode output61

uninstalling Readiris..... 11

user interface 15

user interface language 17

W

windowing templates.....47

Wizard 13

Word output.....61

WordML output.....61

worksheets.....67

X

XML output.....61

XPS iHQC output..... 78

XPS options.....77

XPS output 61, 75