

Chapter 3

COMMAND REFERENCE

This chapter provides an extensive reference guide to the Readiris commands. Note that buttons and shortcuts are available for all frequently used functions!

MAIN TOOLBAR

The main toolbar provides immediate access to the frequently used general commands. Tooltips are available for all buttons.



Auto button

Executes automatic recognition.

The user intervention is limited to a minimum: he initiates the scanning process and saves the recognition result.

Depending on the image source selected with the "Source" or "Scanner" button on the main toolbar, the scanner scans a page automatically or the user is prompted to open an image file. To zone the document, page analysis is executed on the scanned image or the active windowing layout as loaded into memory with the command "Open Layout" under the "File" menu is applied. (Page analysis ignores the graphics on the scanned page.) Next, the character recognition is executed without interactive learning and the user is invited to save the text result or the text result is placed on the clipboard.

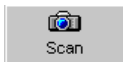
Related Commands

The successive use of the buttons "Scan", "Sort" and "Recognize" on the main toolbar executes step-by-step text recognition. The "Source" and "Scan-



ner" button on the main toolbar determine whether the scanner or prescanned images function as image source.

The "Auto" button is equivalent to the command "Automatic OCR" under the "Process" menu.



Scan button

Scans a document or opens an image file, depending on the selected image source.

The scanned image appears in the image zone progressively.

When the disk option is selected as image source with the "Source" button or when the option "<Image>" is selected as "scanner" with the "Scanner" button on the main toolbar, prescanned images function as image source. Black-and-white TIFF images (uncompressed, packbits, Group 3 and Group 4 compressed), Paintbrush (PCX) images and Windows bitmaps (BMP) are supported. By "scanning", the user is prompted to open a prescanned image file.

Related Commands

The "Scanner" button on the main toolbar determines all scanning parameters such as resolution, page format, brightness and contrast etc.

The rotation buttons on the image toolbar allow the user to rotate images with an incorrect orientation.

The button "Auto" on the main toolbar executes scanning, page analysis and recognition simultaneously.

The "Scan" button is equivalent to the "Acquire" command under the "File" menu and the "Scan" command under the "Process" menu.



Sort button

Allows to modify the sort order of the text windows.

Sorting is very useful to edit the results of the page analysis.

The user clicks inside the windows to define their sequence. The windows he doesn't click on won't be recognized. Arrows indicate the sort order.

The user can display the sort order, type and coordinates of a window by pressing the Alt key while he selects it.

Sorting does *not* apply when "autoformatting" is used - this involves the enabling of the RTF file format and the layout option "Recreate Source Document" with the "Format" button on the main toolbar: "autoformatting" recreates a facsimile copy of the original document and therefore implies that the placement of columns, text blocks and graphics is true to the source document.

Neither does sorting apply to graphic windows: the command "Save Graphics" under the "File" menu saves all graphics in a single file, executing the recognition with "autoformatting" saves the graphics inside the text file.

Related Commands

The "Page Analysis" button on the main toolbar enables the automatic page analysis, the "Analyze Page" button on the image toolbar forces the system to (re-)execute page decomposition.

The button "Auto" on the main toolbar executes scanning, page analysis and recognition simultaneously.

The "Sort" button is equivalent to the "Sort" command under the "Process" menu.



Recognize button

Executes character recognition.

The image file is converted into editable text.

When "autoformatting" is used, Readiris allows to archive a true copy of the user's documents, be it a editable and compact text file instead of a scanned image.



Related Commands

The button "Auto" on the main toolbar executes scanning, page analysis and recognition simultaneously.

The "Recognize" button is equivalent to the command "Recognize" under the "Process" menu.



Language button

Determines the document language.

Indicating the language of the document to be recognized is imperative, as the set of recognized characters varies according to the selected language. Secondly, recognition is assisted by linguistic databases: Readiris extensively uses linguistic information to validate good solutions and mark suspicious ones.

Readiris recognizes all American and European languages, including the Central-European languages, Greek, Turkish, the Cyrillic ("Russian") and Baltic languages. Multi-linguistic support ensures that "exotic" characters such as ç, ß, ñ, γ and ø are recognized correctly.

(It takes the appropriate Windows configuration to display Central-European, Greek, Turkish, Cyrillic and Baltic characters. The user may have to install the Windows "service pack" "Multilanguage Support" before his Windows system is able to cope with these languages. Depending on the software package, the Readiris CD-ROM may contain detailed information on how to install this Windows module.)

The user can press a letter key to move to it directly: if English is currently selected, and he wants to select Hungarian, he can click the "H" key on his keyboard to go to the Hungarian language directly. When several languages have the same initial, he simply presses that letter several times to go through the options. An example: by pressing "P" once, the user selects Polish, by pressing "P" a second time, he selects Portuguese, and by pressing "P" a third time, he's back on Polish.

The British and American - or should we say “international”? - variants of the English language are distinguished.

When the user selects Greek-English or a mixed Cyrillic language setting - for instance "Byelorussian-English" instead of just "Byelorussian" -, a mixed character set is used. Readiris copes beautifully with “Western” words (proper names, brand names etc.) as occur in Greek and Cyrillic documents.

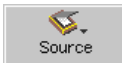
To mix other languages on a page, the user selects the language with the most extended character set. To read documents where English and French occur alongside, French is selected as language to ensure that the accentuated characters such as ç, é and ù are recognized correctly.

To optimally recognize tables of figures, the recognition can be limited to a numeric character set. The symbols 0 to 9, +, *, /, %, , (comma), . (dot), (,), -, =, \$ and £ will then be recognized. This option should not be selected when the tables contain alphabetic text.

Related Commands

The language selection can also be enabled by loading a Readiris configuration with the command "Load Settings" under the "File" menu.

The "Language" button is equivalent to the "Language" command under the "Settings" menu.



Source button

Determines the image source.

The user can capture images using a scanner, or open prescanned images to execute deferred OCR. The latter procedure is useful to recognize faxes.

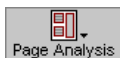
Alternatively, the user can select the option "<Image>" as his “scanner” with the button "Scanner" on the main toolbar to use prescanned images as image source.

The scanner is the default image source.



Related Command

The source selection can also be enabled by loading a Readiris configuration with the command "Load Settings" under the "File" menu.



Page Analysis button

Enables the automatic page analysis.

The user avoids windowing the text zones manually as Readiris decomposes the page automatically. Page analysis is particularly useful when columnized texts and documents with a complex page layout are OCR'd.

Page analysis is fast, skew-tolerant and highly accurate: it traces complex, "irregular" shapes and ignores the graphic zones containing photos, illustrations etc. on the page.

The windows are sorted top-down, left to right to account for columnized documents. Arrows indicate the sort order. The user can change the sort order of windows with the "Sort" button on the main toolbar.

When a table has borders around the cells, page analysis frames it with a single (text) window, and Readiris is able to recreate the table structure.

Although page analysis ignores the graphics, the user can window them manually and include them in an RTF file when "autoformatting" is applied: no sorting applies as "autoformatting" recreates a facsimile copy of the original document, which implies that the placement of columns, text blocks and graphics is true to the source document. Alternatively, the user can save the graphics separately in a single file with the command "Save Graphics" under the "File" menu.

Page analysis is automatically disabled when a windowing template is loaded into memory with the command "Load Layout" under the "File" menu. (The zoning layout remains active until page analysis is re-enabled with the "Page Analysis" button on the main toolbar.)

Related Commands

The command "Analyze Page" on the image toolbar makes Readiris (re-)execute page analysis, the "Sort" button on the main toolbar resorts the text windows.

The page analysis can also be enabled by loading a Readiris configuration with the command "Load Settings" under the "File" menu.

The "Page Analysis" button is equivalent to the option "Page Analysis" under the "Settings" menu.



Learn button

Enables the learning phase, where Readiris acquires new font information.

Font training can substantially enhance the accuracy of the recognition system. When the user tries to read distorted, defaced forms as can be found in real documents or stylized font shapes which Readiris does not recognize optimally, training can overcome this temporary "failure".

User learning is also used to train the system on special symbols which Readiris is unable to recognize, such as mathematical and scientific symbols and dingbats. Readiris can for instance be trained to recognize the " π " symbol as "pi" or the dingbat "☎" as "Tel". (However, the list of recognized symbols cannot be extended with the symbols " π " and "☎"!)

Learning can be disabled when the recognition rate is excellent and the text result requires no detailed proofreading.

Learning is enabled by default. Learning is always disabled with automatic OCR.

The button's tooltip indicates the font dictionary and its mode: the user can hold the mouse cursor over the "Learn" button to learn which user dictionary and dictionary mode are enabled.

Related Commands

The command "Font Dictionary" under the "Learn" menu selects or defines a font dictionary and its mode, the options "New Dictionary", "Append Dictio-



nary" and "Read Dictionary" under the "Learn" menu also set the dictionary mode.

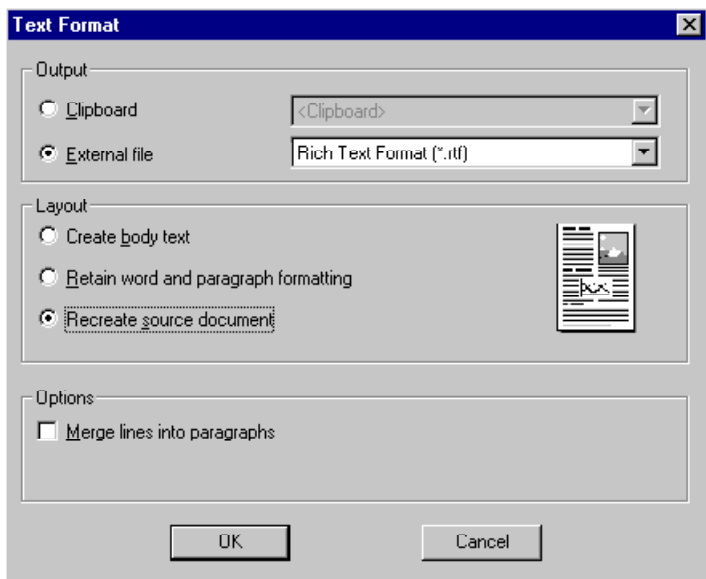
Interactive learning can also be enabled by loading a Readiris configuration with the command "Load Settings" under the "File" menu.

The "Learn" button is equivalent to the option "Interactive Learning" under the "Learn" menu.



Format button

Determines the text file format and the degree of formatting of the recognition result.



The user can select a document format supported by his text applications or copy the output directly to the Windows clipboard. Readiris outputs directly to a large number of wordprocessors and spreadsheets.

This command also determines the level of formatting given to the text output. The availability of formatting options depends on the selected file format: layout reconstruction is limited to the RTF format, other, "poor" formats generating "plain" text such as ASCII, ANSI etc. do not support advanced formatting codes and therefore cannot offer word, paragraph or page formatting.

(RTF is a widely used text format that can be opened by any popular wordprocessor. WordPad is a "reduced" text editor, not a fully featured wordprocessor: WordPad opens RTF files but ignores most formatting elements such as text blocks, alignment etc. To see the recognized text correctly, the user may have to enable the "WYSIWIG" mode of his wordprocessor, mostly called "page layout" mode.)

The user selects "RTF" (Rich Text Format) as output format and "Recreate Source Document" as layout option to apply "autoformatting". Full "autoformatting" means that Readiris recreates a facsimile copy of the scanned document: the word, paragraph and page formatting of the original document are applied. The graphics are included in the RTF file.

Similar typefaces (serif - sans serif, proportional - fixed) are used as in the source document, the point sizes and tpestyles (bold, italic and underlined) are maintained across the recognition. The tabs and the alignment (left - justified) of each text block are recreated. The placement of columns, text blocks and graphics follow the source documents. In other words, Readiris allows to archive a true copy of your documents, be it a editable and compact text files instead of scanned images!

The sorting of windows does *not* apply when "autoformatting" is used: any sorting the user may have done on the results of the page analysis is simply ignored.



The layout option "Create Body Text" creates a non-formatted, "running" text. The text will be captured, but its formatting is entirely ignored. This option is used when a text needs to be recaptured but not its layout.

The option "Retain Word and Paragraph Formatting" represents the middle road: the word formatting - font type, point size and typestyle - is retained across the recognition, and so is the paragraph formatting - the tabs and the alignment. No page formatting is applied: this option just puts one paragraph after the other, it does not recreate columns or copy the relative position of the various text blocks, nor will it include the graphics.

The option "Merge Lines into Paragraphs" sets the text flow within the paragraphs. Enabling this option makes Readiris apply the normal wordwrap typical of wordprocessors until a new paragraph starts, otherwise, a carriage return (CR or EOL code) is added after each line and hyphenated words remain so. Paragraph detection is enabled by default.

Here's an example to clear things up. When the first three lines of a column are "The new presi-", "dent waved from the balcony." and "His wife had joined him.", the paragraph detection yields the following result: "The new **president** waved from the balcony. **His** wife had joined him." The hyphenated parts of the word "president" were "reglued" and a space was added at the end of the first sentence, thus creating naturally flowing text. Had paragraph detection *not* been enabled, the original layout would have been retained, with a carriage return added at the end of each line.

When tables get detected by the page analysis - they have borders around each cell -, the user must select the layout option "Retain Word and Paragraph Formatting" or "Recreate Source Document". The table gets recreated by adding Tab codes between the cells. (The command "Convert Text to Table" of his wordprocessor can be used to round things off.)

When a table has no borders around each cell, the user draw windows around each column to recreate the table structure. In this case, the user must select a table format.

Readiris outputs tables of figures directly to a spreadsheet in the Excel format. The table format creates a generic table file that can be imported directly in the spreadsheet, the clipboard table formats sends the output to the clipboard. The "Paste" command of the spreadsheet is used to paste the recognized table.

To import recognized tables in a Lotus 1-2-3 spreadsheet, the Excel clipboard format is used and the recognized table is pasted in the Lotus worksheet with the "Paste" command, or the Excel format is selected to create a table file, the table is imported in Lotus 1-2-3 and the data are "parsed" afterwards to place the table nicely in cells.

Related Commands

The text format and layout formatting options can also be enabled by loading a Readiris configuration with the command "Load Settings" under the "File" menu.

The "Format" button is equivalent to the "Text Format" command under the "Settings" menu.



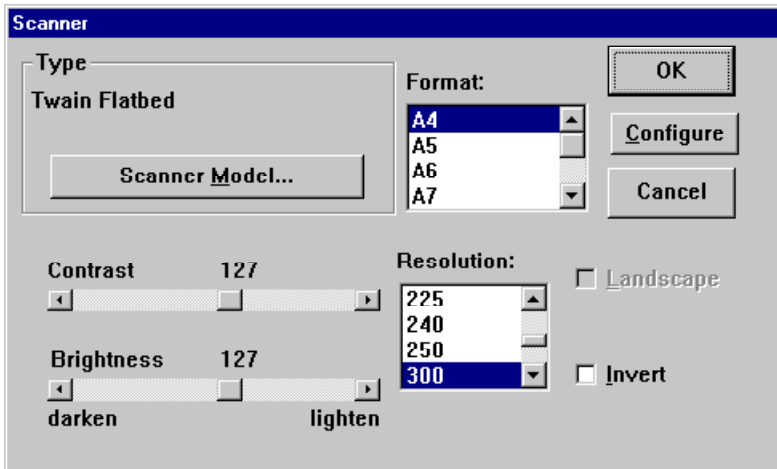
Scanner button

Allows the user to select his scanner model and determine its settings.

The user clicks the button "Scanner Model" to determine his scanner model. If the user's Readiris software licence was bundled with a scanner model, this step is probably unnecessary as his scanner may already be set up under Readiris.

Readiris supports a wide range of popular scanners; it also supports the scanning standard Twain and the scanning platforms Visioneer PaperPort, HP Document Assistant, HP JetSuite, Microtek PageSuite and Primax PaperEase. (Depending on the software bundle, the user may also find specialized manuals concerning the scanning platforms on the Readiris CD-ROM.)

When the user selects the option "<Image>" as his "scanner", prescanned images function as image source.



Once the scanner is selected, this command allows the user to set the scanning resolution, the page format and orientation, brightness and contrast and allows him to indicate whether he is going to use the scanner's document feeder. With Twain compliant scanners, the scanning parameters are often set within the Twain interface.

A resolution of 300 dpi is selected for normal applications, a higher resolution of 400 dpi is recommended for small print (below 10 point) and when the document is very degraded.

By enabling the option "Landscape", the user indicates that the selected page orientation is wide ("landscape") rather than tall ("portrait"). The page orientation actually applies to reduced page formats: on an A4 flatbed scanner, the user can scan A5 pages (half that big) in portrait or landscape format, but he can only scan the full A4 surface in one direction.

The user sets the brightness, and, if available, the contrast. The contrast setting is only available on some scanners. The slider bar is dragged to modify the brightness and contrast; the selected values are represented numerically.

The option "Invert" allows the user to generate “inverted” images - this option is enabled to process full pages with white text on a black background.

The "Configure" button is only available when the scanner allows it. It gives access to some advanced scanning parameters; with Twain scanners, clicking the "Configure" button allows to select the Twain source. Alternatively, the user can use the command "Select Source" under the "File" menu.

Related Commands

The scanner selection and scanner setup can also be enabled by loading a Readiris configuration with the command "Load Settings" under the "File" menu.

The command "Select Source" under the "File" menu allows to select the Twain source.

The "Scanner" button is equivalent to the "Scanner" command under the "Settings" menu.

IMAGE TOOLBAR

The image toolbar contains the tools needed to zone the images comfortably and provides quick access to all frequently used commands required to handle scanned images.

This toolbar is only activated in the “preview” mode, when an image is displayed on-screen. Tooltips are available for all buttons.



Select Window button

Activates the window selection or “arrow” tool.

To modify, move and delete windows, the user has to select them first.

The window selection tool is the default image tool: when an image is displayed on-screen, this tool is enabled automatically.



The user selects the "Select Windows" button on the image toolbar and clicks inside a window. Rectangular markers now appear at each corner and in the middle of the window sides.

To select additional windows, the user holds down the Shift key while clicking on these extra windows.

To select a window and the included windows (of the other type), he holds down the Ctrl key while clicking on the main window.

To unselect windows, he clicks the mouse button elsewhere.

Related Commands

The user can display the sort order, type and coordinates of a window by pressing the Alt key while he selects it. (Graphic windows have no sort order.)

Windows can be cut, copied and deleted with the corresponding commands of the "Edit" menu. Windows can also be modified and moved with mouse cursor movements.



Text Window button

Creates a text window.

To window a text zone of interest, the user selects the "Text Window" tool and clicks the cursor in a corner of the window, stretches the window by moving the mouse to the opposite corner and clicks again.

Sides smaller than 1 mm are not allowed - windows of that size wouldn't contain any characters.

The user can frame "irregular" text blocks by drawing polygonal windows around them. Non-rectangular windows are created by "merging" rectangular zones: as soon as two rectangles (of the same window type) intersect, they become a single window automatically.

The windows are sorted in the order of creation: arrows indicate the sort order. The user can modify the sort order - to include the relevant zones and

exclude the zones of no interest, or to modify their order in the output file - with the "Sort" button on the main toolbar.

The sort order is ignored when "autoformatting" is used - this involves the enabling of the RTF file format and the layout option "Recreate Source Document" with the "Format" button on the main toolbar: "autoformatting" recreates a facsimile copy of the original document and therefore implies that the placement of columns, text blocks and graphics is true to the source document.

Related Commands

The user can display the sort order, type and coordinates of a window by pressing the Alt key while he selects it. The tooltip of the "Text Window" button indicates the number of text windows.

Windows can be cut, copied and deleted with the corresponding commands of the "Edit" menu. Windows can also be modified and moved with mouse cursor movements.

The "Graphic Window" button on the image toolbar allows to create graphic windows.



Graphic Window button

Creates a graphic window.

As the OCR software Readiris generates black-and-white images, no greyscale or color images, it doesn't make sense to scan photographs with Readiris, but the user can capture lineart graphics effortlessly.

To create a rectangle around a illustration, cartoon etc., the user selects the "Graphic Window" tool and clicks the cursor in the upper left corner of the window, stretches the window by moving the mouse to the lower right corner and clicks again.

Sides smaller than 1 mm are not allowed - bitmaps of that size hardly contain any information.



“Irregular”, non-rectangular windows are invisibly converted into rectangles: Readiris covers the maximal surface based on the largest coordinates.

The user can create several graphic windows per operation, but these will be collected, “stacked” in a single file.

Graphics are included directly in the text file when “autoformatting” is applied - this involves the enabling of the RTF file format and the layout option "Recreate Source Document" with the "Format" button on the main toolbar -, and saved separately with the command "Save Graphics" under the "File" menu.

The user can display the type and coordinates of a window by pressing the Alt key while he selects it. (Graphic windows have no sort order.) The tooltip of the "Graphic Window" button indicates the number of graphic windows.

Related Commands

Windows can be cut, copied and deleted with the corresponding commands of the "Edit" menu. Windows can also be modified and moved with mouse cursor movements.

The "Text Window" button on the image toolbar allows to create text windows.



Analyze Page button

(Re-)executes page analysis on the current image.

The user avoids windowing the text zones manually as Readiris decomposes the page automatically. Page analysis is particularly useful when columnized texts and documents with a complex page layout are OCR'd.

Page analysis is fast, skew-tolerant and highly accurate: it traces complex, “irregular” shapes and ignores the graphic zones containing photos, illustrations etc. on the page.

The windows are sorted top-down, left to right to account for columnized documents. Arrows indicate which text windows are included in the output and

their sort order. The user can change the sort order of windows with the "Sort" button on the main toolbar.

When a table has borders around the cells, page analysis frames it with a single (text) window, and Readiris is able to recreate the table structure.

Although page analysis ignores the graphics, the user can window them manually and include them in an RTF file when "autoformatting" is applied: no sorting applies as "autoformatting" recreates a facsimile copy of the original document, which implies that the placement of columns, text blocks and graphics is true to the source document. Alternatively, the user can save the graphics separately in a single file with the command "Save Graphics" under the "File" menu.

Related Commands

The button "Page Analysis" on the image toolbar enables automatic page decomposition, the "Sort" button on the main toolbar resorts the text windows.

The "Analyze Page" button is equivalent to the "Analyze Page" command under the "Process" menu.



Fit to Window button

Displays the entire page.

This zoom level is convenient for seeing the various text and graphic zones in relation to each other.

The user can toggle between "Fit to Window" and "Actual Size" by clicking the *right* mouse button.

Related Commands

The "Fit to Width" button displays the entire width of the document, the "Actual Size" button on the image toolbar and the commands "50% Actual Size" and "200% Actual Size" under the "View" menu display the document at its real size, at half size and at double that size. At actual size, a screen pixel corresponds to an image pixel.



The "Fit to Window" button is equivalent to the "Fit to Window" command under the "View" menu.



Fit to Width button

Displays the entire width of a document.

Related Commands

The "Fit to Window" button on the image toolbar displays the entire document, the "Actual Size" button on the image toolbar and the commands "50% Actual Size" and "200% Actual Size" under the "View" menu display the document at its real size, at half size and at double that size. At actual size, a screen pixel corresponds to an image pixel.

The "Fit to Width" button is equivalent to the "Fit to Width" command under the "View" menu.



Actual Size button

Displays the document at its real size. A screen pixel corresponds to an image pixel.

The user can toggle between "Fit to Window" and "Actual Size" by clicking the *right* mouse button.

Related Commands

The "Fit to Window" button on the image toolbar displays the entire document, the "Fit to Width" button on the image toolbar displays the entire width of a document. The commands "50% Actual Size" and "200% Actual Size" under the "View" menu display the document at half and at double its real size.

The "Actual Size" button is equivalent to the "Actual Size" command under the "View" menu.



Rotate Left button

Rotates the scanned image 90° to the left.

The rotation takes a few seconds, as the image itself is updated, not just the display on-screen. If necessary, the page analysis is re-executed.

Related Commands

The "Rotate Right" button on the image toolbar rotates the scanned image 90° to the right, the "Turn Upside Down" button on the image toolbar rotates the scanned image 180°.

The "Rotate Left" button is equivalent to the "Rotate Left" command under the "View" menu.



Rotate Right button

Rotates the scanned image 90° to the right.

The rotation takes a few seconds, as the image itself is updated, not just the display on-screen. If necessary, the page analysis is re-executed.

Related Commands

The "Rotate Left" button on the image toolbar rotates the scanned image 90° to the left, the "Turn Upside Down" button on the image toolbar rotates the scanned image 180°.

The "Rotate Right" button is equivalent to the "Rotate Right" command under the "View" menu.



Turn Upside Down button

Rotates the scanned image 180°.

This command is used to correct the orientation of a page which was scanned upside down without having to rescan.



The rotation takes a few seconds, as the image itself is updated, not just the display on-screen. If necessary, the page analysis is re-executed.

Related Commands

The "Rotate Left" button on the image toolbar rotates the scanned image 90° to the left, the "Rotate Right" button on the image toolbar rotates the scanned image 90° to the right.

The "Turn Upside Down" button is equivalent to the "Turn Upside Down" command under the "View" menu.

FILE MENU

The "File" menu contains all commands relating to file management: the management of windowing templates and Readiris configurations. It also allows the user to open prescanned images and select a Twain source.

Open

Opens an image file to execute deferred OCR, regardless of the image source selected with the "Source" or "Scanner" buttons on the main toolbar.

Readiris supports black-and-white TIFF images (uncompressed, packbits, Group 3 and Group 4 compressed), Paintbrush (PCX) images and Windows bitmaps (BMP).

Opening prescanned images is particularly useful to convert faxes into editable text files. Should the user have any influence over his correspondents, he can ask them to send faxes with the "fine" quality - those faxes have the higher resolution of 200 dpi and will yield better OCR results.

The user can also use "drag and drop" to open prescanned images. When an image file is dragged from the Windows Explorer onto the Readiris image zone, it is opened promptly.

Finally, the user can select the disk as input source with the "Source" button on the main toolbar and click the "Scan" button, and he can select the option

"<Image>" as "scanner" with the button "Scanner" on the main toolbar and click the "Scan" button to open prescanned images.

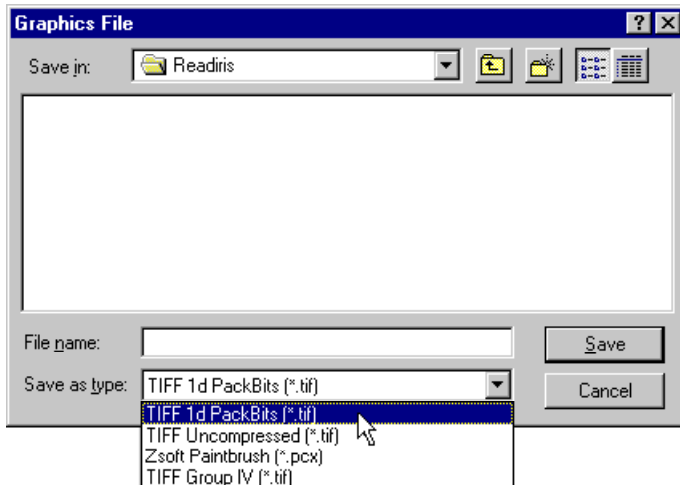
Related Command

The command "Save Graphics" under the "File" menu allows to save images which can be opened and recognized later on.

Save Graphics

Saves the graphic windows.

This command saves all graphics (illustrations, cartoons etc.) zoned with the "Graphic Window" tool in a graphic file (without text recognition).



Graphic zones are never sorted; the user can create several graphic windows per operation, but these will be collected, "stacked" in a single file. (He can use the "Crop" command of his paint or photo retouching program to separate them.)

This command also determines the graphics format in which the graphics will be saved. The user should select a graphic format supported by his paint or



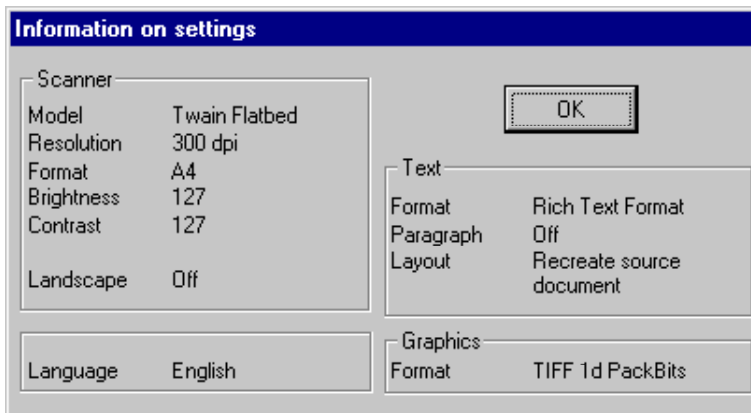
photo retouching software. The file formats TIFF (uncompressed, packbits and Group 4 compressed) and Paintbrush (PCX) are available.

Saving graphics in image files leads to an interesting capability: the user can run deferred OCR on prescanned images. He can scan a document, save the entire page as a graphic zone and recognize the image file later on.

Graphics can also be included directly in the text file when "autoformatting" is applied - this implies the enabling of the RTF file format and the layout option "Recreate Source Document" with the "Format" button on the main toolbar.

Info

Displays the current Readiris configuration.



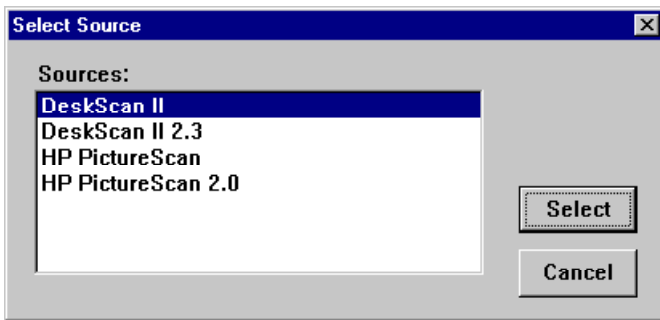
Related Commands

The command "Save Default Settings" under the "File" menu saves the current settings as default settings to avoid having to redefine them each time. Settings are loaded into memory with the command "Load Settings" under the "File" menu and saved with the command "Save Settings" under the "File" menu.

Select Source

Allows to select a Twain source.

This command strictly applies to Twain compliant scanners. Refer to the button "Scanner" on the main toolbar for more information on scanner setup.



Related Command

This command is equivalent to the button "Configure" under the "Scanner" button on the main toolbar.

Acquire

Refer to the "Scan" button on the main toolbar.

Load Layout

Loads a windowing layout into memory.

Layout files contain predefined windowing structures. Zoning templates are particularly useful when documents with a similar page layout are OCR'd, as the user avoids having to redraw the same layout.

Loading a windowing template into memory automatically disables page analysis. The zoning layout remains active until page analysis is re-enabled with the "Page Analysis" button on the main toolbar.



The active layout is applied when documents are automatically recognized with the "Auto" button on the main toolbar.

Related Commands

Zoning layouts can also be enabled by loading a Readiris configuration with the command "Load Settings" under the "File" menu.

The command "Save Layout" under the "File" menu saves the current layout as defined on a scanned document in a windowing template for future use.

Layouts can also be loaded into memory by loading a Readiris configuration with the command "Load Settings" under the "File" menu.

Save Layout

Saves the current zoning layout as drawn on the image in a windowing template.

Layout files contain predefined windowing structures. Zoning templates are particularly useful when documents with a similar page layout are OCR'd, as the user avoids having to redraw the same layout again.

Related Command

The command "Load Layout" under the "File" menu loads a layout into memory.

Load Settings

Loads a Readiris configuration into memory.

Settings files determine which scanner model is being used and its settings, the file format for the output text, the language and characteristics of the document to be recognized etc. *All* operational parameters of Readiris are saved in the settings files.

Related Commands

Settings are saved with the command "Save Settings" under the "File" menu. As the basic settings rarely need to be changed, the command "Save Default

Settings" under the "File" menu saves the current settings as default settings to avoid having to redefine them each time.

Save Settings

Saves the current Readiris configuration in a settings file for later use.

Settings files determine which scanner model is being used and its settings, the file format for the output text, the language and characteristics of the document to be recognized etc. *All* operational parameters of Readiris are saved in the settings files.

Related Commands

As the basic settings rarely need to be changed, the command "Save Default Settings" under the "File" menu saves the current settings as default settings to avoid having to redefine them. Settings are loaded into memory with the command "Load Settings" under the "File" menu.

Save Default Settings

Saves the current Readiris configuration as default settings.

Settings files determine which scanner model is being used and its settings, the file format for the output text, the language and characteristics of the document to be recognized etc. *All* operational parameters of Readiris are saved in the settings files.

The default settings are loaded into memory automatically at each startup. As these settings rarely change, the user avoids having to redefine them.

Related Commands

Settings are saved with the command "Save Settings" under the "File" menu and loaded into memory with the command "Load Settings" under the "File" menu.

Exit

Quits the Readiris software.



Related Command

The default settings saved with the command "Save Default Settings" under the "File" menu will be loaded into memory automatically at the next startup.

EDIT MENU

The commands under the "Edit" menu are limited to the “preview” mode: they are available only when an image is displayed in Readiris' image zone.

They concern the windows: the user can delete, cut, copy and paste them. Shortcuts are available for all commands.

Undo

Cancels the last operation.

This command undoes cut, copy and paste operations of text and graphic windows. If the user erroneously deleted, moved, resized etc. windows, "Undo" will cancel the last operation. This button does not “undo” window selection operations.

Cut

Transfers the selected window(s) to an internal buffer.

Choosing the "Window Selection" or “arrow” tool on the image toolbar and clicking inside a window suffices to select it. Rectangular markers appear at each corner and in the middle of the window sides. To select additional windows, the user holds down the Shift key while he clicks on these extra windows. To select a window and the included windows (of the other type), he holds down the Ctrl key while clicking on the main window.

Related Commands

The command "Paste" under the "Edit" menu allows to insert the selected window(s) stored in an internal buffer. The command "Clear" under the "Edit" menu deletes the selected window(s) irretrievably. The command "Undo" under the "Edit" menu undoes delete operations.

Copy

Copies the selected window(s) to an internal buffer.

Choosing the "Select Windows" or "arrow" tool on the image toolbar and clicking inside a window suffices to select it. Rectangular markers appear at each corner and in the middle of the window sides. To select additional windows, the user holds down the Shift key while he clicks on these extra windows. To select a window and the included windows (of the other type), he holds down the Ctrl key while clicking on the main window.

Related Commands

The command "Paste" under the "Edit" menu allows to insert the selected window(s) stored in an internal buffer. The command "Cut" under the "Edit" menu cuts the selected window(s) to an internal buffer, the command "Clear" under the "Edit" menu deletes the selected window(s) irretrievably.

Paste

Pastes the window(s) stored in an internal buffer.

Windows are pasted in their original position, and the user has to drag them to their new location.

Related Commands

The command "Copy" under the "Edit" menu copies the selected window(s) to an internal buffer. The command "Undo" under the "Edit" menu undoes delete operations.

Clear

Deletes the selected window(s) irretrievably.

Choosing the "Select Windows" or "arrow" tool on the image toolbar and clicking inside a window suffices to select it. Rectangular markers appear at each corner and in the middle of the window sides. To select additional windows, the user holds down the Shift key while he clicks on these extra windows.



To select a window and the included windows (of the other type), he holds down the Ctrl key while clicking on the main window.

Related Commands

The command "Cut" under the "Edit" menu transfers the selected window(s) to an internal buffer. The command "Undo" under the "Edit" menu undoes delete operations.

Select All

Selects all windows of both window types simultaneously.

Related Commands

The command "Cut" under the "Edit" menu transfers the selected window(s) to an internal buffer. The command "Copy" under the "Edit" menu copies the selected window(s) to an internal buffer. The command "Paste" under the "Edit" menu allows to insert the selected window(s) stored in an internal buffer. The command "Clear" under the "Edit" menu deletes the selected window(s) irretrievably. The command "Undo" under the "Edit" menu does not “undo” selection operations.

SETTINGS MENU

The "Settings" menu allows to specify the document characteristics and OCR options. It also allows to configure the scanner.

Scanner

Refer to the "Scanner" button on the main toolbar.

Language

Refer to the "Language" button on the main toolbar.

Font Type

Determines the font type of the document to be recognized.



The setting "Normal" is used for the printing qualities typeset, typewritten, laser printed, inkjet printed and "letter quality" dot matrix printing (also called "25 pin" or "NLQ" dot matrix), the setting "Dot matrix" is strictly reserved for draft dot matrix printing (also called "9 pin" dot matrix); such texts require highly specialized recognition techniques.

Related Commands

The font type can also be enabled by loading a Readiris configuration with the command "Load Settings" under the "File" menu.

Character Pitch

Determines the character pitch of the document to be recognized.



The setting "Fixed" is selected when the document is composed of a fixed font. With fixed or "monospaced" typefaces, each character of the typeface has the same width. An "i" takes up as much horizontal space on a line as a "w", as is the case in this sentence. Such typefaces are typically found in documents produced by typewriters.

The setting "Proportional" is selected when the document is composed of a proportional font. With proportional fonts, the width of the letters depends on their shape. A "w" takes up more horizontal space on a line than an "i", as is the case in this sentence. Virtually all fonts found in newspapers, magazines and books are proportional.

The setting "Auto" is selected to have Readiris detect the character pitch automatically. It is enabled by default and can be left untouched at all times.



Related Command

The character pitch can also be enabled by loading a Readiris configuration with the command "Load Settings" under the "File" menu.

Text Format

Refer to the "Format" button on the main toolbar.

View after Scan

Enables the "preview" mode, where the scanned documents are displayed on-screen in the image zone.

All commands concerning the image preview - the buttons on the image toolbar and the commands under the "Edit" menu - are available only when an image is opened in the image zone.

This minor option is enabled by default, and is best left untouched.

Page Analysis

Refer to the "Page Analysis" button on the main toolbar.

VIEW MENU

The "View" menu contains the zoom levels and rotation commands. A minor diagnostic command is added. Buttons on the image toolbar provide quick access to all frequently used commands.

Fit to Window

Refer to the "Fit to Window" button on the image toolbar.

Fit to Width

Refer to the "Fit to Width" button on the image toolbar.

50% Actual Size

Displays the document at half its real size. (At actual size, a screen pixel corresponds to an image pixel.)

Related Commands

The "Fit to Window" button on the image toolbar displays the entire document, the "Fit to Width" button on the image toolbar displays the entire width of a document. The commands "Actual Size" and "200% Actual Size" under the "View" menu display the document at actual size and at double its real size.

Actual Size

Refer to the "Actual Size" button on the image toolbar.

200% Actual Size

Displays the document at twice its real size. (At actual size, a screen pixel corresponds to an image pixel.)

Related Commands

The "Fit to Window" button on the image toolbar displays the entire document, the "Fit to Width" button on the image toolbar displays the entire width of a document. The commands "50% Actual Size" and "Actual Size" under the "View" menu display the document at half its actual size and at double its real size.

Rotate Left

Refer to the "Rotate Left" button on the image toolbar.

Rotate Right

Refer to the "Rotate Right" button on the image toolbar.

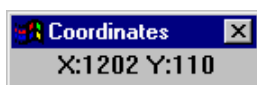
Turn Upside Down

Refer to the "Turn Upside Down" button on the image toolbar.



Coordinates

This minor diagnostic option displays the coordinates of the mouse cursor in document pixels.



Related Command

The user can display the type, coordinates and sort order of a window by pressing the Alt key while he selects it.

PROCESS MENU

The "Process" menu contains the commands needed to scan, decompose, sort and recognize documents.

Automatic OCR

Refer to the "Auto" button on the main toolbar.

Scan

Refer to the "Scan" button on the main toolbar.

Sort

Refer to the "Sort" button on the main toolbar.

Recognize

Refer to the "Recognize" button on the main toolbar.

Analyze Page

Refer to the "Analyze Page" button on the image toolbar.

LEARN MENU

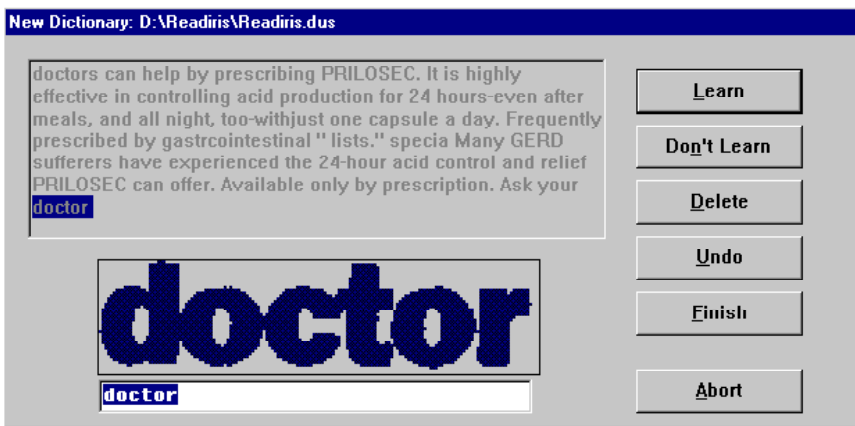
The "Learn" menu contains all commands relating to the interactive learning.

Font training can substantially enhance the accuracy of the recognition system. When the user tries to read defaced and stylized font shapes which Readiris does not recognize optimally, training can overcome this temporary "failure".

Linguistics offer useful help to solve ambiguous cases such as an "O" which might be mistaken for a '0'. Another typical example is the letter "l" and number '1' which have an identical form in many fonts - think of texts produced on old typewriters! The linguistic context helps to determine whether you are dealing with "l" or '1'.

User learning is also used to train the system on special symbols which Readiris is unable to recognize, such as mathematical and scientific symbols and dingbats. Readiris can for instance be trained to recognize the " π " symbol as "pi" or the dingbat "☎" as "Tel". (However, the list of recognized symbols cannot be extended with the symbols " π " and "☎".)

Readiris enters the interactive learning phase at the end of the recognition: the recognized text is displayed progressively and the system stops on doubtful characters, or - in the case of touching characters ("ligatures") - on doubtful character strings. They are always presented in their context, the doubtful characters are highlighted. Unrecognized characters are represented by a tilde (the "~" symbol).



First of all, the user should verify if he has activated the correct font dictionary and dictionary mode - these are always indicated in the title of the learning window (as in the tooltip of the "Learn" button on the main toolbar). If that is not the case, he should click the "Abort" button - the last image is redisplayed with the zoning as was created -, enable the right font dictionary or dictionary mode and run the OCR again.

If necessary, a character (or character string) is entered for the incorrect or unknown shape and one of the following buttons is clicked.

Learn

The user agrees with the proposed solution or corrects it. The program saves this doubtful character in the font dictionary as "sure", final. Future recognition will no longer require his intervention, the shape is considered learnt once and for all.

In the example above, the system stops on six joined characters, and "Learn" is to accept a shape which cannot be confused with other characters.

Don't Learn

The user agrees with the proposed solution or corrects it. The difference with the "Learn" button is that the learnt symbol gets the status "unsure" in the font dictionary. For future recognition, the system will propose the learnt solution but still require a confirmation.

This button is used for symbols which might be confused with others: a de-faced "e" which might be mistaken for a "c", a damaged "t" which closely resembles an "r" etc.



The "t" above is close to the "r" symbol -, and the user clicks "Don't Learn" so as not to confuse the two symbols further on.

Delete

The displayed form is eliminated from the output. This button is used to ignore "noise" on the documents - spots, coffee stains etc. - which might get recognized as points, comma's etc. -, and to erase any other unwanted symbol.

Undo

The user goes back to correct mistakes. He can undo the nine last decisions.



Finish

The learning process is aborted but the OCR continues in automatic mode. All decisions by the system thereafter are accepted without user validation.

This button is clicked when the recognition is highly accurate and does not require detailed proofreading.

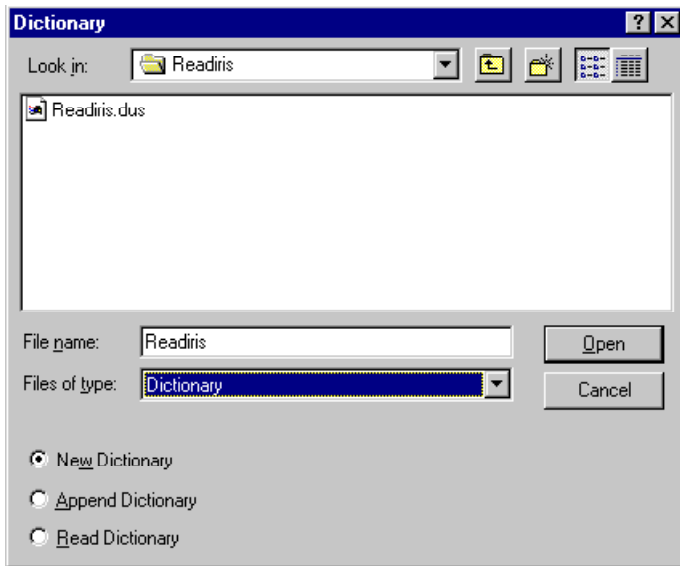
The "Finish" button should not be confused with the "Abort" button: with the "Abort" button, no output is generated and the OCR must be run again, with the "Finish" button, the text is created, it just isn't proofread in detail.

The results of a training session are stored in files called "font dictionaries" for future use. These font dictionaries can be loaded to make good use of previous training, so that Readiris will recognize the learnt character shapes automatically.

Font Dictionary

Selects or defines a font dictionary and its mode of operation.

User dictionaries contain font information the system learns during the interactive learning. When many documents of the same type are recognized, the user can avoid having to go through the same learning process every time. The learnt characters can be stored in a font dictionary and that information can be used when documents of the same type are recognized.



The user also sets the dictionary mode: "New Dictionary" creates a new dictionary, the user starts without any previous knowledge. The option "Append Dictionary" uses the previously learnt font information and completes it, the option "Read Dictionary" uses the previously learnt font information but does not complete it further.

The new mode is selected when a single page is recognized. To recognize many pages of the same type - pages with the same fonts and printing quality - the new mode is selected for the first page, the append mode for a few pages more and the read mode for the rest of the document(s). When multipage documents are recognized, scanning a next page automatically puts the font dictionary in the append mode.

As font dictionaries are limited to 500 learnt shapes, the user is recommended to create specific dictionaries for specific applications, for instance per type of



document. For clarity, it is recommend to give meaningful names to the font dictionaries, for instance REPORT.DUS, PALATINO.DUS *dus* etc. Dictionaries have the default extension *.DUS.

When the user enters a filename for a new dictionary and activates the “append“ mode, an empty font dictionary is created to be completed.

When a dictionary is full, the learning has no further effect: the results of the learning are no longer held in memory or written to a dictionary.

Related Commands

The dictionary mode can also be set with the options "New Font Dictionary", "Append Font Dictionary" and "Read Font Dictionary" under the "Learn" menu.

The tooltip of the "Learn" button on the main toolbar indicates the font dictionary and its mode.

The "Learn" button on the main toolbar enables the interactive learning.

A font dictionary and its mode can also be enabled by loading a Readiris configuration with the command "Load Settings" under the "File" menu.

New, Append, Read Font Dictionary

Sets the mode of the font dictionary.

"New Font Dictionary" creates a new dictionary, the user starts without any previous knowledge. The option "Append Font Dictionary" uses the previously learnt font information and completes it, the option "Read Font Dictionary" uses the previously learnt font information but does not complete it further.

The new mode is selected when a single page is recognized. To recognize many pages of the same type - pages with the same fonts and printing quality - the new mode is selected for the first page, the append mode for a few more pages and the read mode for the rest of the document(s). When multipage documents are recognized, scanning a next page automatically puts the font dictionary in the append mode.

When a dictionary is full, the learning has no further effect: the results of the learning are no longer held in memory or written to a dictionary.

The tooltip of the "Learn" button indicates the font dictionary and its mode.

Related Commands

The mode can also be set with the options "New Dictionary", "Append Dictionary" and "Read Dictionary" of the "Font Dictionary" command under the "Learn" menu, and can be enabled by loading a Readiris configuration with the command "Load Settings" under the "File" menu.

The "Learn" button on the main toolbar enables the interactive learning, the command "Font Dictionary" under the "Learn" menu selects a font dictionary and sets its mode.

Interactive Learning

Refer to the "Learn" button on the main toolbar.

HELP MENU

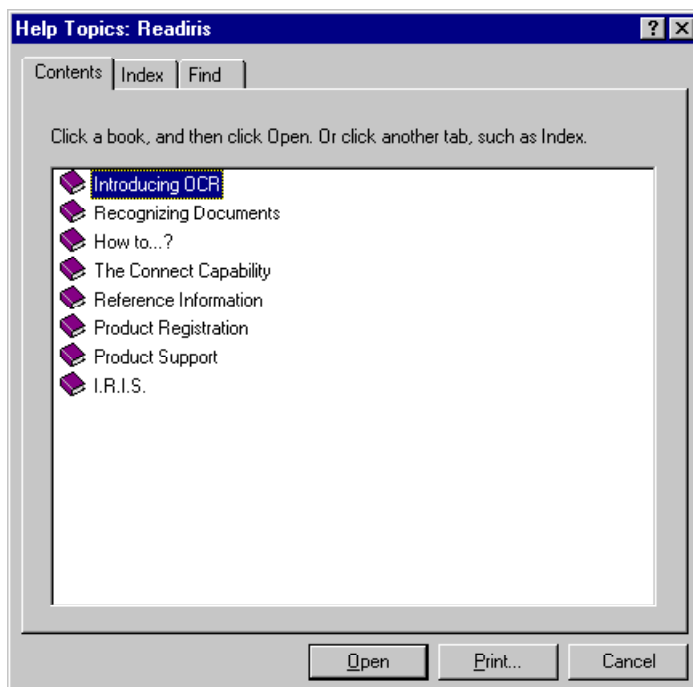
The "Help" menu contains the on-line help system.

Some last-minute information may not be included in this manual. The user is recommended to consult the on-line help system for additional information on Readiris' capabilities.

Help Topics

Allows to navigate through the on-line help system.

The available information cannot only be consulted on-screen but also printed out.

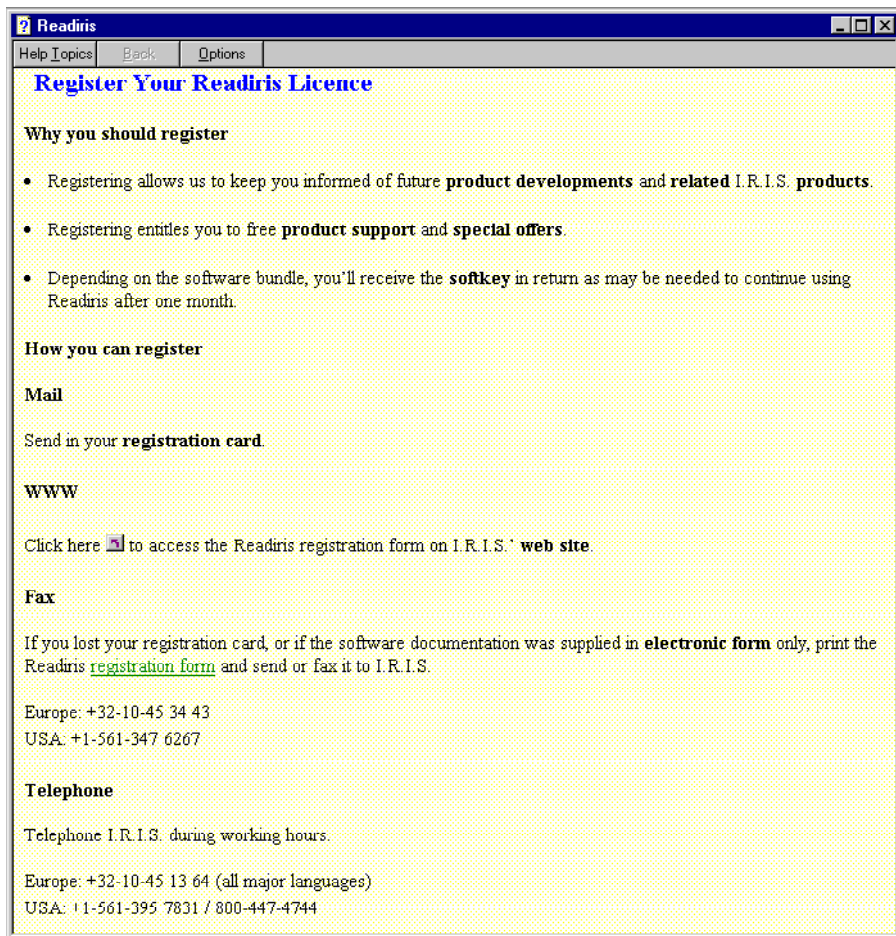


Register

Informs the user how to register his Readiris software licence.

The user can register in many ways: by sending in the registration card or faxing its electronic counterpart, by calling I.R.I.S. during working hours and by filling out a registration form on the I.R.I.S. web site.

Depending on the software bundle, he'll receive the softkey in return as may be needed to continue using the Readiris software after one month.





Registering the Readiris licence will allow I.R.I.S. to keep the user informed of future product developments and related I.R.I.S. products. The registration benefits, including free product support and special offers, are strictly limited to registered users.

Product Support

Informs the user how to get Readiris product support.



Contact I.R.I.S.

Gives the coordinates of the I.R.I.S. offices.





About Readiris

Displays the startup screen where the software version and copyrights are displayed.

The user clicks anywhere to return to the Readiris software.