

How the OCR Process Works

- The document is read by your scanner. The scanner acts as the “eye” of your computer and sends it the image in a digitized form. At this point, the **scanned image** is no more than a meaningless cloud of black points (pixels) on a white background.
- The OCR software extracts **text** information from these pixels: it recognizes the shapes and assigns characters.
- **Line segmentation** consists of slicing a page of text into its different lines. This step also analyses **lineskew**, interline spacing and drop letters, and separates **touching lines**.
- The **word segmentation** isolate one word from another.
- The **character segmentation** separates the various letters of a word. If the characters have the same width (fixed pitch), this step is easy. The problem gets more interesting when the width of the letters depends on their shape (proportional pitch), when kerning occurs and when dot matrix fonts are used.
- The **character recognition** extracts characteristics out of each isolated shape and assigns a symbol. The three most important stages are the autolearning phase, topological analysis and the optional interactive phase. During the recognition, linguistic knowledge is used to validate correct solutions and flag suspicious ones.

Character Isolation

- Readiris copes beautifully with **touching characters** (“ligatures”) and Kerning.
- For **dot matrix** fonts, highly specific segmentation techniques are used.
- Recognition of any font is possible if the system manages to isolate a sufficient number of characters. Readiris does *not* read extremely **stylized**, “script”-like **fonts**, where character segmentation is impossible.

Character Size

- Readiris reads fonts with a **size range** of 6 to 72 points (1” high).

6 point

72 point

- Readiris copes with drop letters.

Tip: select a resolution higher than 300 dpi to read **small print** (smaller than 10 point).

Printing Quality

- Readiris reads virtually any **printing quality**: typeset, typewritten, laser printed, inkjet printed, letter quality dot matrix printed texts and draft dot matrix printing.
- **Draft dot matrix** printed texts - where characters are composed of clouds of isolated dots - require highly specific segmentation and recognition techniques. Indicate the correct font type.
- All **handwritten** texts are excluded.

Related Topic

Font Type

Character Sets

- Readiris recognizes all **American** and **European** character sets, including the **Central-European** languages and the **Greek** and **Cyrillic** (“Russian”) alphabets.
- Optionally, you can read **Japanese** and Simplified **Chinese**.
- The list of **recognized characters** varies according to the selected document language.
- When you select **Cyrillic-English**, **Greek-English**, **Japanese** or Simplified **Chinese**, a **mixed character set** is used. Readiris copes beautifully with “Western” words (proper names etc.) as occur in Cyrillic, Greek and Asian documents.
- Readiris can limit recognition to a **numeric character set** to recognize tables of figures which contain no text.
- Readiris can be trained on any **special symbol** - be it mathematic or scientific symbols, dingbats etc.

How to OCR a Document

- Select the **input source**: you capture images using your scanner or open prescanned images (for instance faxes).
- Be sure that your **scanner** is selected and that its **settings** are correct.
- **Scan** the page with your scanner or open the image file.
- Have a closer look at the scanned page to control the **image quality**.
- If the **page analysis** is enabled, the text windows, graphics and tables are detected automatically. Perform “sorting” to select the windows of interest and define their order. If the page analysis is disabled, use manual **windowing** or a zoning **layout** to indicate which blocks you want to recognize.
- Specify the **document's** language and characteristics.
- Start the character **recognition**.
- You may **interrupt** at any time by clicking the Stop button (or pressing Escape).
- If **learning** is enabled, you'll enter the interactive phase.
- **Save** the recognized text in a text file, send it directly to a target application or copy it to the clipboard. You can create a new file or append an existing text file.
- You may be prompted to read **another page**. By answering yes, you scan another page, by answering no, the previous image is redisplayed.

Tip: the **OCR wizard** guides you through the OCR process comfortably - answer a few simple questions and you'll obtain quick and easy results.

Tip: “**automatic**” **OCR** automates the recognition to the max - the user simply initiates the scanning and saves the text result.

Automatic Recognition

“**Automatic**” OCR offers “single-click” OCR. The recognition is reduced to a minimum: the user simply starts the scanning and saves the text result.

- [Page analysis](#) or a [windowing template](#) is used to **decompose** the page.
- Interactive **learning** is disabled.
- All other **settings** apply: the document language, font type, text format etc. are as currently selected.

(The [OCR wizard](#) offers a *semi*-automatic way of recognizing pages.)

Related Topic

[How to OCR a document](#)

How to OCR Tables

Readiris outputs tabular data to **spreadsheets** and **wordprocessors**: tables gets reconstructed cell by cell in worksheets and inserted as table objects in wordprocessor files.

Windowing

- The page analysis detects “**gridded**” and “**non-gridded**” tables: “gridded” or “framed” tables have borders around the cells, “non-gridded” tables are not surrounded by cell borders.
- When the **space** between the columns of “non-gridded” tables is too big, Readiris may avoid windowing them as tables to avoid confusion with text columns.

Selecting the “language”

- When you limit recognition to the **numeric character set**, the symbols 0 to 9, +, *, /, %, , (comma), . (dot), (,), -, =, \$ and £ will be recognized.
- Do not select this option when your tables contain **text**!

Output mode

- To interface with **spreadsheets**, **send** the table directly to Microsoft **Excel** or select a **table format**.
- To interface with **wordprocessors**, **send** the table directly to Microsoft **Word** or select the **text format** Word (DOC) or RTF. You must retain the word and paragraph formatting or recreate the source document.

Using the “Connect” capability

Use the Connect feature to scan tables directly from within the Excel spreadsheet.

How to OCR Faxes

Operation

Readiris opens black-and-white images in the (compressed) TIFF, Paintbrush (PCX) and Windows bitmap (BMP) **format**.

How to...?

Select the disk as **input source** to open prescanned images. Proceed as usual.

Tip: **drag** an image from the Windows Explorer onto the Readiris window.

Tip: should you have any influence over your correspondents, ask them to send faxes with the “**fine**” **quality** - those faxes have the higher resolution of 200 dpi and will yield better OCR results.

Windowing Tools

Operation

- **Windows** are drawn on a displayed image to indicate which **text** and **table zones** you want to recognize and which **graphics** you want to save.
- There are three **window types**: text blocks, graphic zones (containing photos, illustrations etc.) and tables. A color code indicates the window type: text windows are yellow, graphics are blue and table windows are purple.
- There are three different **methods** of zoning images: automatic **page analysis**, **manual windowing** and the use of fixed **layouts**.
- Drawing or detecting several windows is mandatory whenever a text is arranged in **columns**.

Page analysis

- To decompose all scanned documents automatically, enable the **option** Page Analysis.
- To make Readiris decompose the current page, use the **command** Analyze Page.
- To **select** windows and modify their **sort order**, use the Sort command.
- Select the document language *before* executing the page analysis when you are dealing with **Asian documents**. Specific routines are used for these languages: the interline spacing of Asian documents is in most cases bigger than in Western documents, the text is made up of small icons (“ideograms”) that could easily be seen as graphic zones in Western documents and the text may run from top to bottom, from right to left.

Manual windowing

Tools are available to draw **windows** manually on the scanned image.

Page analysis with manual windowing

- Apply **page analysis** to detect the windows and **sort** the ones you want to include.
- Start drawing a window **manually** - all windows you didn't select are erased.

Windows layouts

Specific **zoning templates** can be stored for future use.

Creating Windows

- Select the **text**, **graphic** or **table** window tool. The mouse pointer assumes the shape of a crossbar.
- Move the cursor to a corner of the window you want to create and click.
- Stretch the window by moving the cursor to the opposite corner and click again.

Tip: create **polygonal windows** by merging rectangular zones. Whenever two windows (of the same type) intersect, they become a polygon automatically. (Creating polygonal table windows makes no sense.)

Selecting Windows

- Select the “arrow” tool and click inside a window to **select** a window. Markers appear at each corner and in the middle of the window sides.
- To **unselect** windows, click the mouse button elsewhere.
- To select **additional windows**, hold down the **SHIFT** key while clicking inside these windows.
- To select a window and the **included windows** (of the other type), hold down the **CTRL** key while clicking inside the main window.
- Choose the command Select All to select **all windows**.

Tip: you can display the type, coordinates and sort order of a window by pressing the **ALT** key while you select it!

Modifying Windows

- **Select** the window.
- **Place** your cursor over a marker.
- **Drag** the mouse to change the window size.

Related Topic

[Selecting Windows](#)

Moving Windows

- **Select** the window(s).
- **Drag** the mouse to modify the position of the window(s).

Related Topic

[Selecting Windows](#)

Deleting Windows

- **Select** the window(s). (Choose the command Select All to select all windows.)
- Select the command Cut to **cut** the selected window(s) to an internal buffer.
- Select the command Clear to **delete** the selected window(s) irretrievably.

Tip: use the corresponding **shortcut** keys!

Related Topic

[Selecting Windows](#)

File Menu

Allows to save and retrieve Readiris settings and windowing layouts and opens prescanned images.

Open

Opens (TIFF, PCX and BMP format) image files, mostly faxes.

Tip: drag an image from the Windows Explorer onto the Readiris window.

Save Graphics

Saves the graphic windows.

Save Full Page as Image

Saves the entire page as a graphic for deferred OCR.

Info

Displays the current Readiris settings.

Select Source

Selects and sets a scanner input source. Twain scanner standard feature.

Acquire

Acquires an image. Twain scanner standard feature.

Load Layout

Retrieves a windows layout.

Save Layout

Saves a windowing layout as drawn on an image.

Load Settings

Recovers a Readiris configuration.

Save Settings

Saves the Readiris settings.

Save Default Settings

Saves the current Readiris settings as default for the next startup.

Exit

Quits the Readiris software.

Save Graphics - Save Full Page As Image

Operation

- You can save the **graphic windows** only - without performing recognition on the text and table windows.
- You can **include** the graphics in the text file by selecting the [autoformatting](#) option Recreate Source Document and enabling the option Include Graphics.
- You can save **entire** scanned **pages** in image files for **deferred recognition**.

Saving graphics

- The **command** Save Graphics saves the **graphic windows** only - without performing text recognition on the text windows.
- You can create **several** **graphic windows** per operation, but these will be collected in a **single file**.
- Select a **graphic format** supported by your photo retouching software.
- Select the **graphic window** tool to zone a graphic, illustration etc.

Saving full pages as images - deferred OCR - fax recognition

- The **command** Save Full Page as Image saves the entire page in an image file.
- Do so to run OCR on **prescanned images**. You can scan a document, save the page as a image and recognize the image file later on!
- A similar procedure is used to recognize **faxes**: [faxes](#) in the (compressed) TIFF, Paintbrush (PCX) and Windows bitmap (BMP) format can be opened.

Layouts

Operation

- Layout files contain predefined windowing structures. Zoning templates are particularly useful when documents with a similar **page layout** are OCRed.
- Using layouts is an **alternative** to [page analysis](#) and [manual windowing](#).

How to...?

- Layouts as defined on a scanned document are **saved** with the command Save Layout.
- Layouts are **loaded** with the command Load Layout. Loading a layout into memory disables the page analysis. The layout remains activated until you re-enable page analysis.

Settings

Settings determine **all** operational **parameters** of Readiris: document language, scanner model and settings, output mode and formatting options etc.

How to...?

- As settings don't change at every turn, they can be stored as **default settings** with the command Save Default Settings.
- Settings are **saved** with the command Save Settings.
- Settings are **loaded** with the command Load Settings.

Edit Menu

Operation

- These commands concern the text, graphic and table **windows**. You can delete, cut, copy and paste windows!
- Handy **shortcut** keys are available for all commands.

Selecting windows

- To **select** a window, choose the arrow tool and click inside a window. Markers appear at each corner and in the middle of the window sides.
- To select **additional windows**, hold down the **SHIFT** key while you click inside the window. To select a window and the **included windows** (of the other type), hold down the **CTRL** key while clicking inside the main window.
- Select the command Select All to select **all windows**.
- To **unselect** windows, click the mouse button elsewhere.

How to...?

- Select the command Copy to **copy** the selected window(s) into an internal buffer.
- Select the command Paste to **paste** the contents of the internal buffer. The inserted windows are inserted in their original location. Drag them to another location.
- Select the command Cut to **cut** the selected window(s) to an internal buffer.
- Select the command Clear to **delete** the selected window(s) irretrievably.
- Select the command Select All to select all windows.
- Select the command Undo to **cancel** the previous command.

Related Topic

[Windowing Tools](#)

Scanner

Scanner model

- Select your **scanner** model - the scanner has to be connected to your PC and ready!
- Readiris supports almost every **flatbed**, **sheetfed** and **hand-held scanner**, “all-in-one” **MFPs** (“Multifunctional Peripherals”) and scanner standard currently available. Readiris is **Twain** compliant.
- You’ll find information on scanner setup in the “**Read Me**” files. Refer to these files if you experience any problem with your scanner, and consult the hardware documentation that came with your scanner.
- If you select “<Image>” as “scanner” model, prescanned images function as image source.

Scanner settings

- Set the scanner **settings**: resolution, page format and orientation, brightness, contrast and use of the autofeeder.
- **Options** not supported by the selected scanner will be not displayed. With **Twain** compliant scanners, the scan settings are often set within the Twain module.
- With **HP** scanners, you can enable the option “Auto-Exposure” to automatically adjust the brightness and contrast to the scanned document.
- 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.

Language

- Readiris recognizes all **American** and **European** languages, including the **Central-European** languages, **Greek** and the **Cyrillic** (“Russian”) languages.
- Optionally, you can read **Japanese** and Simplified **Chinese**.
- The list of **recognized characters** varies according to the selected document language.
- Recognition is assisted by a **linguistic database**. Readiris extensively uses linguistic information to validate good solutions and mark suspicious ones.
- When you select **Cyrillic-English**, **Greek-English**, **Japanese** or Simplified **Chinese**, a **mixed character set** is used. Readiris copes beautifully with “Western” words (proper names etc.) as occur in Russian, Greek and Asian documents.
- To **mix** other **languages** on a page, select the language with the most extended character set. To read documents where English and French occur alongside, select French as language to ensure that the accentuated characters - such as ç, é and ù - recognized correctly.
- Limit recognition to a **numeric character set** to optimally OCR **tables of figures**. The symbols **0** to **9**, **+**, *****, **/**, **%**, **,** (comma), **.** (dot), **(**, **)**, **-**, **=**, **\$** and **£** will be recognized. Do not select this option when your tables contain text!
- Select the document language *before* executing the **page analysis** when you are dealing with **Asian documents**. Specific page analysis routines are used for Asian documents.

Related Topic

[Settings](#)

Recognizing Asian Documents

Asian languages

- Readiris [optionally](#) offers recognition of the Asian languages **Japanese** and Simplified **Chinese**. Traditional Chinese is not supported.
- **Simplified Chinese** is used on China's mainland, where Traditional Chinese is used by the “Big 5” communities - Hong Kong, Taiwan, Singapore etc.

Working environment

- To view and edit Asian documents, you can install an Asian, Japanese or Chinese version of the **Windows** operating system.
- Alternatively, you can run specialized “**emulating**” **software** (such as UnionWay AsianSuite or TwinBridge AsianBridge) on a Western version of Windows to correctly represent the ideograms of these languages.

Page analysis

- Select the document language *before* executing the [page analysis](#) when you are dealing with Asian documents.
- Specific routines are used for these languages: the interline spacing of Asian documents is in most cases bigger than in Western documents, the text is made up of small icons (“ideograms”) that could easily be seen as graphic zones in Western documents and the text may run from top to bottom, from right to left.

Character sets

When you select Japanese or Simplified Chinese, a **mixed character set** is used. Readiris copes beautifully with “Western” words (proper names etc.) as occur in Asian documents.

Document characteristics

The commands [Font Type](#) and [Character Pitch](#) do not apply to Asian documents.

Interactive learning

- [Interactive learning](#) does not apply to Asian documents.
- Learning does not make sense for these languages which use thousands of different symbols - and you'd have to be able to enter the ideograms, not an easy task when using a Western keyboard!

Text formatting

- The number of **file formats** is reduced.
- Generating Asian text implies creating **body text** - you cannot apply “autoformatting” or retain the

word and paragraph formatting.

Font Type

- Use the setting **normal** for the printing qualities typeset, typewritten, laser printed, inkjet printed and letter quality dot matrix printing.
- Use the setting **dot matrix** strictly for draft dot matrix printing: such texts require highly specialized segmentation and recognition techniques.
- This command does not apply to Asian documents.

Tip: the **tooltip** of the Recognize button indicates the reading mode. When no message is added, the “normal” font quality applies, when the message "Dot matrix" is added, the dot matrix reading mode is enabled.

Related Topics

Character Isolation

Settings

Character Pitch

- Select the **fixed pitch** if the characters have the same [width](#).
- Select the **proportional pitch** if the [width](#) of the letters depends on their shape.
- This command does not apply to Asian documents.

Tip: select **auto** to have Readiris detect the character pitch for you.

Related Topic

[Settings](#)

Text Format

Output modes

- Select an output mode: save the results to an **external file** (with a document format supported by your text application), place the results in the Windows **clipboard** or **send** the recognized document **to a target application**.
- Readiris outputs to a large number of **wordprocessors**, **spreadsheets** and “poor” text formats that generate “plain text”. Readiris exports data directly to Microsoft **Word**, Microsoft **Excel** and the Windows accessory **WordPad**.

Text formatting

- The option **Create Body Text** avoids text formatting by Readiris: you get a continuous, running text. All formatting, if any, is done afterwards by the user.
- With the option **Retain Word and Paragraph Formatting**, the font type (serif - sans serif, proportional - fixed, normal - condensed), size and typestyle (bold - italic - underlined) are maintained across the recognition. The tabs and the alignment (left - centered - right - justified) of each block are recreated. However, no graphics are captured, the text blocks and columns aren’t recreated - the paragraph just follow each other. The tables are recaptured correctly.
- The option **Recreate Source Document** recreates a **facsimile copy** of the original document: the text blocks, tables and graphics are recreated in the same place and the word and paragraph formatting are maintained across the recognition. As a result, you get a true copy of your source document, be it a compact and editable text file, no longer a scanned image of your document.
- Which formatting options are **available** depends on the selected **output mode** and **language**: the options Retain Word and Paragraph Formatting and Recreate Source Document are limited to the Word (DOC) and RTF formats. “Poor” text formats generating “plain” text (such as ANSI) do not support advanced formatting codes. For **Asian documents**, output is limited to body text.

Options

- **Merge Lines into Paragraphs** enables the automatic **paragraph** detection. This prevents the insertion of carriage returns (CR or EOL codes) at the end of each line, Readiris wordwraps the recognized text until a new paragraph starts, and “reglues” hyphenated words at the end of a line.
- **Include Graphics** includes the [graphics](#) in “autoformatted” text files.

Table recognition - spreadsheets

Readiris outputs tabular data to **spreadsheets** and **wordprocessors**: tables gets reconstructed cell by cell in worksheets and inserted as table objects in wordprocessor files.

- Tables can be sent to the spreadsheet Microsoft Excel and the wordprocessor Microsoft Word, can be sent to the clipboard, saved inside a Word (DOC) and RTF file, and saved using a table format.
- Retain the word and paragraph formatting or recreate the source document.

Asian documents

- The number of **file formats** is reduced.
- Generating Asian text implies creating **body text** - you cannot apply “autoformatting” or retain the word and paragraph formatting.

Related Topic

[Settings](#)

View after Scan

Minor option. Enables the preview mode, where the scanned documents are displayed on-screen. Some commands such as the image toolbar are only available in the preview mode!

Page Analysis

- You avoid the manual windowing of the various zones as Readiris **decomposes the page** automatically. This is particularly useful when columnized texts and documents with a complex page layout - possibly including graphics and tables - are OCR'ed.
- There are three **window types**: text blocks, tables and graphic zones containing photos, illustrations etc. A color code indicates the window type: text windows are yellow, graphics are blue and table windows are purple.
- The windows are **sorted** top-down, left to right to account for **columnized documents** but you can change their sort order.
- Page analysis is fast, skew-tolerant and highly **accurate** - it “detects” irregular shapes.
- The page analysis detects “**gridded**” and “**non-gridded**” tables: “gridded” or “framed” tables have borders around the cells, “non-gridded” tables are not surrounded by cell borders. (When the **space** between the columns of “non-gridded” tables, Readiris may avoid windowing it as a table to avoid confusion with text columns.)
- Page analysis detects zones where you get **white text on a black background**. (Readiris “inverts” the image when the need arises to recognize such text blocks!)

Tip: use the command Analyze Page to make Readiris (re-)execute page analysis.

Tip: select the document language *before* executing the page analysis when you are dealing with **Asian documents**. (Special routines will be used: the interline spacing is in most cases bigger than in Western documents, the text is made up of small icons (“ideograms”) that could easily be seen as graphic zones in Western documents and the text may run from top to bottom, from right to left.)

Related Topic

[Windowing Tools](#)

Enable Wizard on Toolbar

- When this option is enabled, you'll find the OCR Wizard button on the main toolbar, when it is disabled, the same button acts as Auto button to trigger [automatic recognition](#).
- The [OCR wizard](#) offers a **semi-automatic** way of recognizing pages: it guides you through the OCR process comfortably. Answer a few simple questions and you'll obtain quick and easy results.

Tip: you can make the OCR wizard start running each time you **start up** Readiris with the option [Enable Wizard on Startup](#). (It is enabled by default.)

Enable Wizard on Startup

Operation

- When this option is enabled, the [OCR wizard](#) starts running running each time you **start up** Readiris.
- The OCR wizard offers a **semi-automatic** way of recognizing pages: it guides you through the OCR process comfortably. Answer a few simple questions and you'll obtain quick and easy results.

How to...?

- This option is enabled by **default**.
- The option Enable Wizard on Startup in the first screen of the wizard has the same effect.

Tip: you can place the OCR wizard on the main toolbar with the option [Enable Wizard on Toolbar](#).

View Menu

Is available only when an image is displayed on-screen.

Zoom levels

Zoom levels are useful to verify the scanning quality and to window the scanned image comfortably.

- **Fit to Window** displays the document entirely. You see the various text zones and graphics in relation to each other.
- **Fit to Width** displays the entire width of the document.
- **50% Actual Size**, **Actual Size** and **200% Actual Size** 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.

Tip: toggle between Actual Size and Fit to Window by clicking the right **mouse** button.

Rotation

Rotate Left, **Rotate Right** and **Turn Upside Down** rotate the document 90° to the left, 90° to the right and 180°.

Coordinates

Minor diagnostic option. Displays the coordinates of the mouse cursor in document pixels.

Controlling the Image Quality

The recognition rate is greatly dependant on the document and scanner quality. Taking a closer look at the scanned image is useful to control this quality.

- With too **light documents**, characters get broken up: "O" may become "()" etc.
- Too **dark documents** contain very heavy shapes, open letters get closed and too many characters are glued together: "c" becomes an "o" etc.

How to...?

- Adjust the quality of the scanned image by modifying the **brightness** setting of your scanner. (The document is **rescanned** promptly.)
- **HP** scanners allow to enable the “Auto-Exposure” option, which detects the correct brightness automatically.

Related Topics

Character Isolation

Scanner

OCR Wizard

Operation

- The OCR wizard offers a **semi-automatic** way of recognizing pages. (The [Auto button](#) offers a *fully* automatic way of recognizing pages.)
- The wizard **guides** you through the OCR process comfortably: answer a few simple questions and you'll obtain quick and easy results.

How to...?

- The OCR wizard is enabled on the main **toolbar** with the option [Enable Wizard on Toolbar](#). When this option is enabled, you'll find the OCR Wizard button on the toolbar, when it is disabled, the same button acts as Auto button to trigger [automatic recognition](#).
- The OCR wizard starts running each time you **start up** Readiris when the option [Enable Wizard on Startup](#) is enabled. (This option is enabled by default.)

Scan - Open

- **Scans** a document or opens a prescanned image file, depending on your input source.
- Readiris **opens** (compressed) TIFF, Paintbrush (PCX) and Windows bitmap (BMP) **images**. Prescanned images are mostly faxes.
- Pay some attention to **lineskew**. Although page analysis and recognition are skew-tolerant, it may become difficult to window and OCR a page correctly when the skew is too significant.
- If necessary **rotate** the image with the rotation tools.

Tip: to open prescanned images, **drag** them from the Windows Explorer onto the Readiris window.

Sort

Sorting windows

- Arrows indicate which text windows are included in the output and their **sort order**.
- Sorting is very useful to **correct** the results of the page analysis.
- **Click** inside a window to define their sequence.
- The windows you don't click on won't be **recognized**!
- You can force the system to (re-)execute **page decomposition** with the command Analyze Page.

Page analysis with manual windowing

- Apply **page analysis** to detect the windows and **sort** the ones you want to include.
- Start drawing windows **manually** - all windows you didn't select are erased.

Tip: you can display the **sort order**, type and coordinates of a window by pressing the [ALT](#) key while you select it!

Recognize

- **Starts** the character recognition.
- Click the Stop button (or press Escape) to **interrupt**.
- If **learning** is enabled, you'll enter the [interactive phase](#). (Learning does not apply to Asian documents.)
- Send the recognized text to a target application, copy it to the clipboard or save it in a text file. You can create a new file or append an existing text file.
- If you have enabled [autoformatting](#) with the option Recreate Source Document, the **graphic windows** may be saved *inside* the wordprocessor file. Otherwise, use the command Save Graphics to [save the graphic zones](#) only.
- If the page analysis detected **tables** and the word and paragraph formatting are retained or the source document is recreated, they get recreated as table objects inside the wordprocessor file.
- You may be prompted to read **another page**. By answering yes, you scan another page, by answering no, the previous image is redisplayed.

Tip: the **tooltip** of the Recognize button indicates the reading mode. When no message is added, the “normal” font quality applies, when the message "Dot matrix" is added, the **dot matrix** [reading mode](#) is enabled.

Tip: scanning another page puts the **font dictionary** in the append mode.

Related Topic

[How OCR works](#)

Analyze Page

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

Related Topic

[Windowing Tools](#)

Font Dictionary

Operation

Font dictionaries contain font information the system learns during the interactive learning.

When you recognize many documents of the same type, you may not want to repeat the same learning process every time. Store the learnt characters in a dictionary and use that information when you recognize documents of the same type.

Commands

- **Select** or define a font dictionary.
- Don't forget to set the appropriate **dictionary mode**: New Dictionary creates a new dictionary, you start without any previous knowledge. Append Dictionary uses the previously learnt font information and completes it. Read Dictionary uses the previously learnt font information but does not complete it further.

Tip: if you have only **one page** to be recognized, select the new mode. If you have **many pages** of the same type - pages with the same font and printing quality - select the new mode for the first page, the append mode for a few pages more and the read mode for the rest of your document(s).

Tip: with **multipage documents**, scanning a next page automatically puts the font dictionary in the append mode.

Related Topics

[Interactive Learning](#)

[Learning Fonts](#)

New, Append or Read Dictionary

Select the appropriate mode for the font dictionary:

- **New Dictionary** creates a new dictionary. You start without any previous knowledge.
- **Append Dictionary** uses the previously learnt font information and completes it.
- **Read Dictionary** uses the previously learnt font information but does not complete it further.

Tip: if you have only **one page** to be recognized, select the new mode. If you have **many pages** of the same type - pages with the same font and printing quality - select the new mode for the first page, the append mode for a few pages more and the read mode for the rest of your document(s).

Tip: with **multipage documents**, scanning a next page automatically puts the font dictionary in the append mode.

Related Topics

[Interactive Learning](#)

[Learning Fonts](#)

Interactive Learning

- Enables the learning phase, where Readiris acquires new **font** information.
- Disable learning when the **recognition rate** is excellent.
- Learning is always disabled with **automatic OCR** and when **Asian documents** are read.

Related Topics

[Learning Fonts](#)

Learning Fonts

If Readiris is unsure about a reading result, the doubtful character (or character string) is displayed with its parent word and the proposed solution is presented.

Verify (in the window title) if the correct **font dictionary** and dictionary mode are enabled - click the button Abort and start again if this is not the case.

Enter the valid character (or character string) if necessary and select one of the following buttons.

Learn

You agree with or correct the proposed solution. Readiris saves the solution in the dictionary as **sure**. Future recognition will no longer require your intervention.

Don't Learn

You agree with or correct the proposed solution. Readiris saves the solution in the dictionary as **unsure**. Future recognition will still require your intervention.

Use this command for damaged characters which could be confused with other characters if learned. A typical example is the characters '**l**' and '**I**', which have an identical form in many fonts.

Delete

Deletes characters from the output. Use this button to prevent **noise** on your document from appearing in the output file as periods, commas etc.

Undo

You go back, correct mistakes. Readiris keeps track of the nine last operations.

Finish

Aborts the learning phase, Readiris runs **automatically** thereafter. You accept every solution of the software.

Abort


Aborts the OCR. This option is used when an incorrect **font dictionary** or dictionary mode are enabled. All results are deleted and you start over again. (Don't confuse with Finish, where the recognition results are maintained!)

Related Topics

[Interactive Learning](#)

Register Menu

Registration Wizard

Click here  to start the registration wizard.

Purchase Info

Gives information on how to purchase a software licence from I.R.I.S.

Register

Gives access to [registration info](#) as may be needed to get a **softkey** from I.R.I.S.


Register Your Readiris Licence

Why you should register

- Registering allows us to keep you informed of future **product developments** and **related I.R.I.S. products**.
- Registering entitles you to free **product support** and **special offers**.
- Depending on the software bundle, you'll receive the **softkey** in return as may be needed to continue using Readiris after one month.

How to...?


Registration wizard

Click here  to start the registration wizard.

Mail

Send in your **registration card**.

WWW

Click here  to access the Readiris registration form on I.R.I.S.' **web site**.

Fax

If you lost your registration card, or if the software documentation was supplied in **electronic form** only, print the Readiris [registration form](#) and send or fax it to I.R.I.S.

Europe: +32-10-45 34 43

USA: +1-561-347 6267

Telephone

Telephone I.R.I.S. during working hours.

Europe: +32-10-45 13 64 (all major languages)

USA: +1-561-395 7831 / 800-447-4744

Readiris Registration Form

Serial number licence

Identification number licence

Company

Department

Title

Address

Zip Code / City

State / Country

Telephone

Fax

E-mail

Scanner model

Place of purchase

Purchase date

Date

Image Recognition Integrated Systems
Product Registration
Rue du Bosquet 10, 1348 Louvain-la-Neuve (Belgium)
Tel: 32-10-45 13 64 - Fax: 32-10-45 34 43
1600 N.W. Boca Raton Bld. Suite 20
Boca Raton, FL 33432 (USA)
Tel: 1-561-395 7831 / 800-447-4744 - Fax: 1-561-347 6267
Support: support@irislink.com

Help Menu

Help Topics

Provides general information on OCR and reference information on Readiris.

Product Support

Informs you about the [product support](#) offered by I.R.I.S.

Contact I.R.I.S.

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

About Readiris

Displays the Readiris startup screen with the copyright notices. Click to make it disappear.

How to Get Product Support

Free technical support is offered to all **registered customers**. ([Registering](#) also entitles you to special offers.)

Europe

Hotline: +32-10-45 13 64 (working hours) (all major languages)

Fax: +32-10-45 34 43


USA

Hotline: +1-561-395 7831 (working hours)

Fax: +1-561-347 6267

WWW

www.irislink.com/support.html (troubleshooting info)

Click here  to access the troubleshooting info.

E-mail

support@irislink.com

How to Get in Touch with I.R.I.S.

Head Office (Belgium)

Phone: +32-10-45 13 64


Fax: +32-10-45 34 43

I.R.I.S. on the Internet

Web site: <http://www.irislink.com>

E-mail sales: sales@irislink.com

E-mail support: support@irislink.com

Click here  to access the I.R.I.S. home page.

USA Office (East Coast)

Phone: +1-561-395 7831 / 800-447-4744

Fax: +1-561-347 6267

USA Office (West Coast)

Phone: +1-480-854-3111 / 800-7USAIRIS

Fax: +1-480-854-2929

France Office

Phone: +33-1-646 16 762

Fax: +33-1-646 16 071

Software Option

Reading Asian documents

The software option “**Asian OCR add-on**” offers recognition of the Asian languages **Japanese** and **Simplified Chinese**. Traditional Chinese is not supported. (Simplified Chinese is used on China's mainland, where Traditional Chinese is used by the “Big 5” communities - Hong Kong, Taiwan, Singapore etc.)

What it takes - the working environment

To view and edit Asian documents, you can install an Asian, Japanese or Chinese version of the Windows operating system or run specialized “emulating” software (such as UnionWay AsianSuite or TwinBridge AsianBridge) on a Western version of Windows to correctly represent the ideograms of these Asian languages.

Acquiring it

[Contact](#) I.R.I.S. to obtain this software option.

Installing the Connect Capability

Preliminary step

Start Readiris, select your scanner and save the [default settings](#).

Word 97 (Office 97)

- Start Word. Open the document [CONWRD97.DOT](#).
- Create a new file.

The command Scan Text is added to the File menu.

Word 2.0 to 7.0 (Office 95)

- Start Word. Open the document [CONNECT.DOC](#) in the Readiris folder for Word 2.0 and Word 6.0, or open the document [CONNECT7.DOC](#) for Word 7.0 (Office 95).
- Answer the questions posed by the system. Answer yes when you are asked to save the global command changes.
- Create a new file.

The command Scan Text is added to the File menu.

Excel 97 (Office 97)

Start Excel. Open the document [CONX97.XLS](#) in the Readiris folder.

The menu Scan Table is added to the menu bar.

Excel 3.0 to 7.0 (Office 95)

- Copy the file [CONNECT3.XLM](#) if you are running Excel 3.0, [CONNECT4.XLM](#) if you are running Excel 4.0 or 5.0, [CONNECT7.XLM](#) if you are running Excel 7.0 (Office 95), from the Readiris folder to the subfolder [XLSTART](#) of Excel.
- Rename the copied file [CONNECT.XLM](#).
- Start Excel.

The command Scan Table is added to the File menu.

WordPerfect 5.x and 6.0

- Start WordPerfect.
- Install the Connect macro under WordPerfect. Your WordPerfect manual can tell you how: with

WordPerfect 5.x, the command Assign to Menu under the Macro menu is used, with WordPerfect 6.0, the command Macro under the Tools menu.

- Enter the macro name [CONNECT.WCM](#) for WordPerfect 5.x or [CONNECT6.WCM](#) for WordPerfect 6.0, preceded by the pathname of the Readiris folder.

The new command is added to the Macro menu.

Uninstalling a Connect Macro

Items to remove

It takes two manual steps to “uninstall” a Connect macro: you have to remove the macro, and remove the added command from your command menus.

Removing the macro

- Remove the macro **irisocr**.
- How you remove macros differs for each application. Consult the documentation that came with your software program to learn how to remove a macro. Generally speaking, you should look for a **Macro** command.

Restoring the original menu commands

- Remove the menu **command** Scan Text or Scan Table from the command menu or remove the **menu** Scan Table from the menu bar.
- How you remove a menu or menu commands differs for each application. Consult the documentation that came with your software program to learn how to remove a macro. Generally speaking, you should look for a **Customize** command.

Using the Connect Capability

- Readiris can *not* run in the background.
- Click on the **new command** in your application to run Readiris.
- The Connect interface may ask you some **questions** before Readiris starts running.
- Proceed as usual and quit the Connect interface when you are through. The recognition results are inserted at the current **cursor position**.

brightness

The balance of light and dark colors in an image. Scanner function or button to lighten and darken the image data.

compression

Reduction of the amount of storage space required to save an image. Compressed images are saved more efficiently, resulting in smaller files. However, they take longer to open because they must be decompressed first.

contrast

The relative difference between the dark and light portions of an image. Scanner function or button to lighten and darken the image data.

draft dot matrix printing

Texts printed on a matrix printer where you can still distinguish the individual dots that compose a symbol. Also called “9 pin” dot matrix.

ape~descended l i f e

drop letter

Large capital letter that covers several lines of text. “Dropped” caps are used to emphasize the beginning of a paragraph or section.

As a quick way to publish and access information, there is nothing quite like the Internet. In just a couple of years, it's transformed the way most companies

dpi

Short for “dots per inch”. Unit of measure that indicates the image resolution in number of dots per linear inch. The higher the value, the higher the resolution.

fixed font

With fixed fonts, each character of the typeface has the same width. An "i" takes up as much horizontal space on a line as a "w". Also called “monospaced” fonts. Documents produced with old typewriters have a fixed pitch.

image

The bitmapped result of scanning a document. Using hardware, the document is converted into a collection of dots stored in a file that can be processed by the computer.

kerning

In printed texts, some letter pairs have more space between them than others because of their shape and slate. Pair-kerning automatically reduces the space between such letter pairs to enhance their appearance. Otherwise, they would be spaced too close or too far apart to be aesthetically pleasing.



The image shows two versions of the word "WALTER" in a serif font. The top version, "WALTER", has a wide space between the 'A' and 'L' and a narrow space between the 'T' and 'E'. The bottom version, "WALTER", has a narrow space between the 'A' and 'L' and a wide space between the 'T' and 'E'. The letters 'W' and 'E' are red, 'A' and 'L' are blue, and 'T' and 'R' are black.

While the first "WA" and "LT" are non-kerned letter pairs, the second "WA" and "LT" are clearly kerned.

letter quality dot matrix printing

Dot matrix printing where you can no longer distinguish the individual dots that compose a symbol. The printing quality approaches the quality of a laser printer. Also called “NLQ” or “25 pin”.

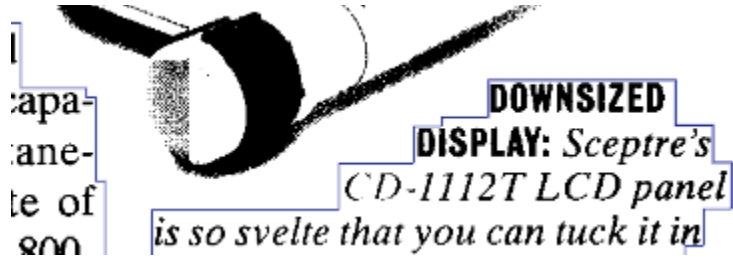
marker

Small rectangle at the window corners and in the middle of the window sides. Markers indicate that the window is selected.

page analysis

Detection and sorting of the text blocks as occur on a scanned page. Graphic zones containing photos, artwork etc. are ignored. Page decomposition is particularly useful when columnized texts and documents with a complex page layout are OCRed.

Page analysis is skew-tolerant and highly accurate - it detects irregular shapes.

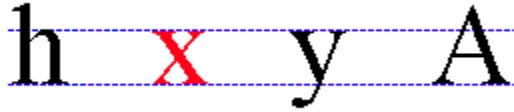


pixel

Short for “picture element”. The smallest distinguishable square dot to compose a bitmapped picture. The higher the resolution, the smaller the pixels.

point

Typographical unit of measure that describes the font size. A point equals 1/72 inch. Putting things the other way around: a point corresponds to 0.014 inch (0.35 mm).



The height of the lowercase "x" of a typeface serves as base for the measurement.

preview

Mode of operation where the scanned image is displayed on-screen.

proportional font

Proportional fonts allow the characters of the typeface to have a different width. A "w" will take up more horizontal space on a line than an "i". Virtually all fonts you find in newspapers, magazines and books are proportional.

resolution

A measure of the fineness of detail that the scanner produces. The number of pixels of an image file is expressed in dots per linear inch. The higher the resolution, the sharper the image.

scanner

Electronic input device designed to digitize images. Converts physical pictures into bitmapped graphic files that can be processed by a computer.

scanner standard

International communication protocol used by several scanner manufacturers to standardize interfacing between software and scanner. Twain is an example.

topological analysis

Readiris analyzes the segmented characters as human beings unconsciously do: a number of features - strokes, loops, holes, nodes, angles etc. - are extracted and checked against a predefined or learnt source of knowledge.



Topological analysis is to a large extent independant from the character size.

