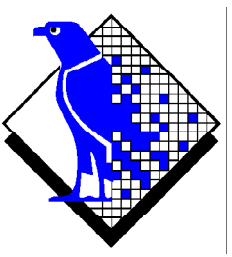
Visioneer PaperPort

Readiris, Cardiris and ScanSoft PaperPort

Using Readiris and Cardiris with the Visioneer/ScanSoft PaperPort scanning platform



Readiris, Cardiris and the Visioneer/ScanSoft PaperPort software

This short manual discusses how to use the OCR software Readiris and business card organizer Cardiris with the Visioneer-ScanSoft PaperPort 3.5 and the 4.0 to 5.2 “Deluxe” scanning software. (The differences between PaperPort 3.5 and the later “Deluxe” versions are pointed out where they are relevant.)

It contains all you need to know to get started with these I.R.I.S. products successfully using the PaperPort scanning platform - for instance on a Visioneer or HP scanner. However, this manual does not intend to replace the PaperPort, Readiris or Cardiris User's Guide. To make full use of the many advanced capabilities of these software packages, it is mandatory that you study their comprehensive user documentation closely.

INSTALLING I.R.I.S. PRODUCTS UNDER PAPERPORT

The Readiris software is delivered exclusively on an autorunning CD-ROM. To install, simply insert the CD-ROM in your CD-ROM drive and wait for the installation program to start running. Follow the on-screen instructions.

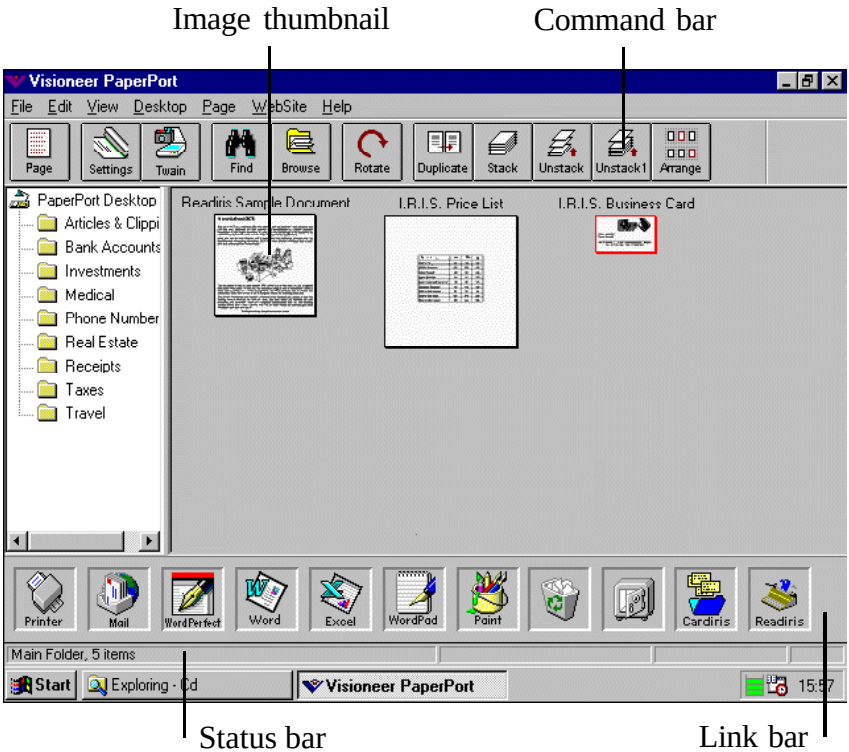
Are any additional steps required to install Readiris and Cardiris under PaperPort? Should you set some parameters from within the PaperPort scanning platform? The answer is simply no. As soon as Readiris and Cardiris are installed correctly, the **PaperPort link** is established automatically.

It doesn't even matter whether you install Readiris and Cardiris *before* or *after* you install the PaperPort software! Simply install these software packages on your PC and you can be up and running immediately.

STARTING UP THE PAPERPORT SCANNING SOFTWARE

Let's have a closer look at this. Start the PaperPort software by double-clicking the corresponding icon. (If you are equipped with a sheetfed or keyboard scanner, you can also push the button on the left side of your scanner or insert a paper document in the scanner to start the application.)

When the PaperPort desktop is displayed, you will notice that the Readiris and Cardiris icon are included in the **link bar**.



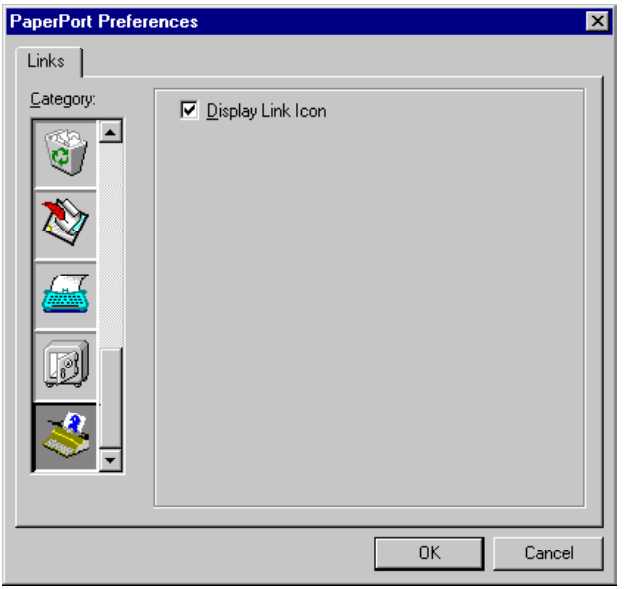
"Drag and drop" is how you do things with the scanned images: to OCR a *text* or a *table* of figures, it suffices to drag the image thumbnail onto the Readiris icon. In this way, you indicate that you want to make use of the scanned document as a text. To archive *business cards*, simply drop the images onto the Cardiris icon. (You can also select an item and just click on the link icon.)

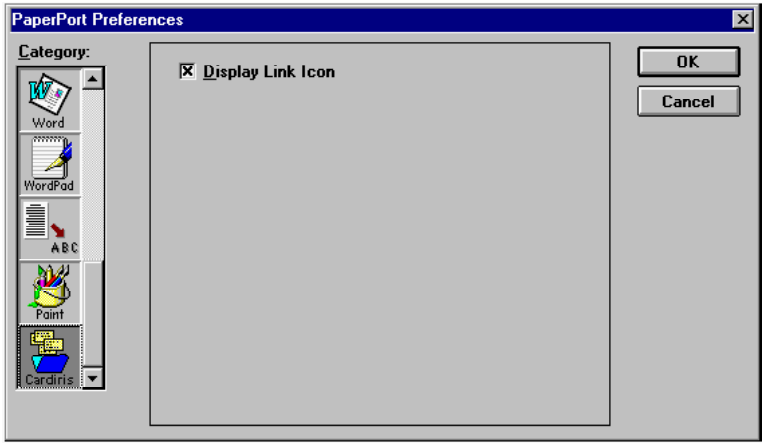
With PaperPort 3.5, the Readiris OCR function is actually represented several times on the desktop. With that version of the PaperPort software, Readiris is not only available as an icon on the link bar but is also "hidden" behind your wordprocessor and spreadsheet. You can drop a table of figures onto the spread-

sheet icon to OCR it. As OCR converts images into editable text, your wordprocessor and spreadsheet take care of text recognition as well!

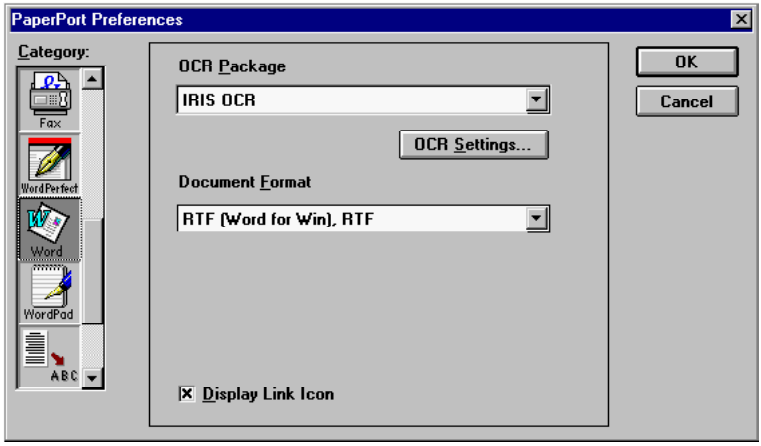
Click the "Settings" button on the command bar to see how Readiris and Cardiris were set up under PaperPort.

For Readiris and Cardiris, things are very simple. When you select the *Readiris* or *Cardiris* link with the command "Link Preferences" under the "Edit" menu (PaperPort “Deluxe” versions) or when you select the Readiris or Cardiris icon in the "Category" list (PaperPort 3.5), a single option is displayed, "Display Link Icon". This option is enabled by default; leave it activated at all times, otherwise the Readiris and Cardiris application are no longer available on the link bar.



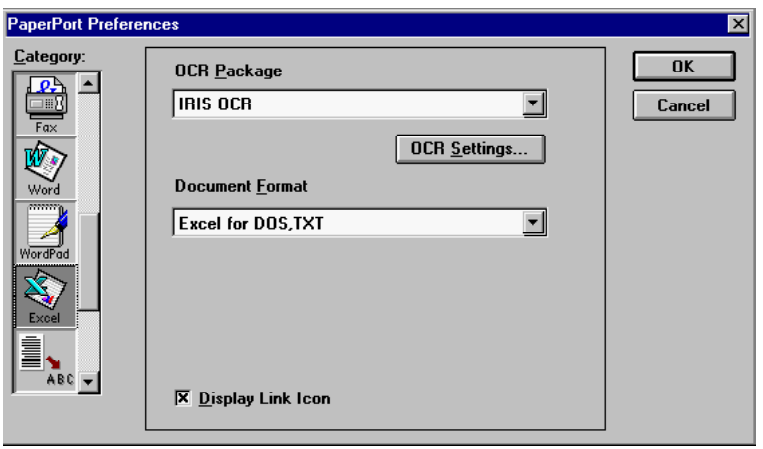


As the OCR function is also available under the wordprocessor and spreadsheet icon with PaperPort 3.5, you can select your wordprocessor icon to see how Readiris is configured under PaperPort.



Readiris has been installed automatically as the OCR tool linked to your *wordprocessor* and the RTF text format will be used by default to transfer texts to your wordprocessor. You can select another file format, but this is again of no importance: the file format is only used internally by Readiris to transfer the text result to your wordprocessor. We recommend you not to modify the default parameters.

The same holds for your *spreadsheet*: the OCR link with Readiris is established automatically and the correct file format is enabled by default.

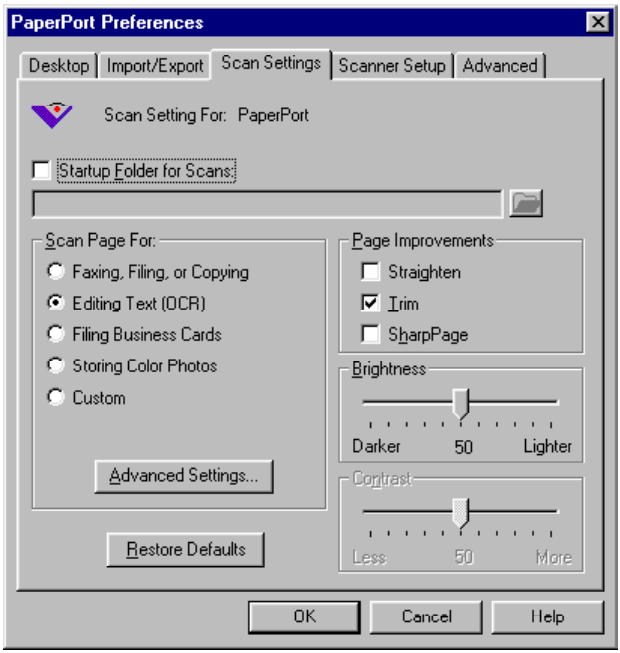


SETTING YOUR SCANNER FOR READIRIS

Summing up: to OCR a document, you scan it from the PaperPort desktop and drag its thumbnail onto the Readiris icon. If you are using PaperPort 3.5, you can also drop the image on the wordprocessor or spreadsheet icon.

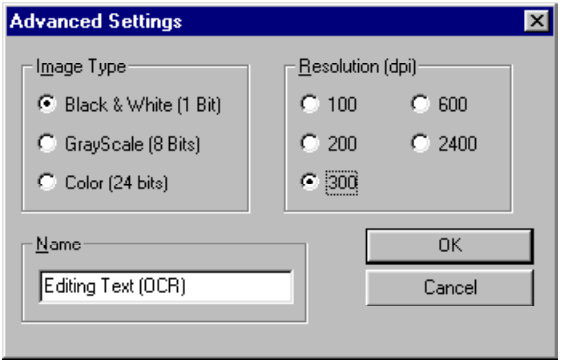
Be sure that you scan your document with the correct settings. Click the "Settings" button on the command bar to verify this. The scanning settings are displayed by default.

Things are somewhat different for PaperPort 3.5 and the “Deluxe” versions. Let's start with PaperPort “Deluxe”. Select the **scan mode** "Editing Text (OCR)". This is the standard setting for OCR purposes.

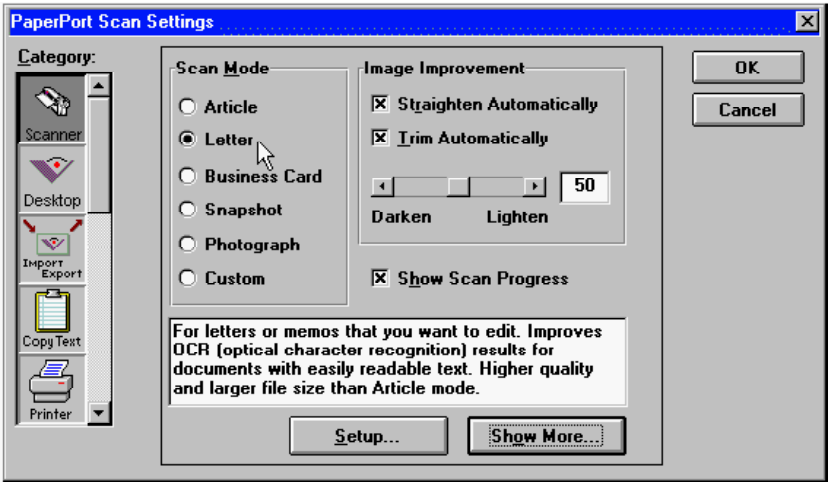


What these settings imply becomes clear when you click the button "Advanced Settings": black & white images are scanned at a resolution of 300 dpi.

With version 4.0, SharpPage image enhancement will be applied and skewed pages will be straightened automatically to optimize the images for the consecutive OCR process. With version 5.2, the white border surrounding the scanned image is erases as the image gets “trimmed”. Don't hesitate the modify these settings!

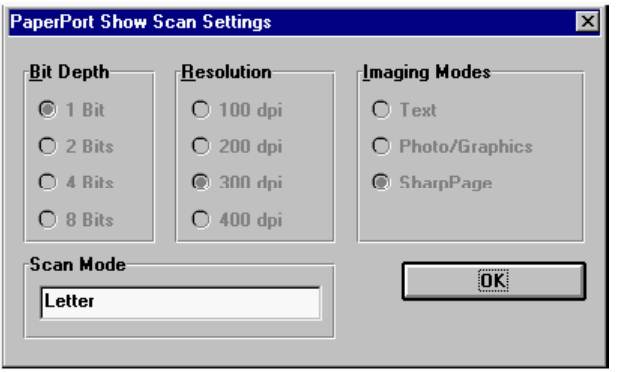


On to PaperPort 3.5. Here, you click the "Settings" button on the command bar and select the **scan mode** "Letter".



Click the button "Show More" to see what these settings stand for: black & white images are scanned at a resolution of 300 dpi with application of SharpPage

image enhancement to optimize the images. You are recommended to automatically apply image deskewing with the option "Straighten Automatically" and image cropping with the option "Trim Automatically".



To know which scan settings were used for a specific image, select the image and check the status bar.

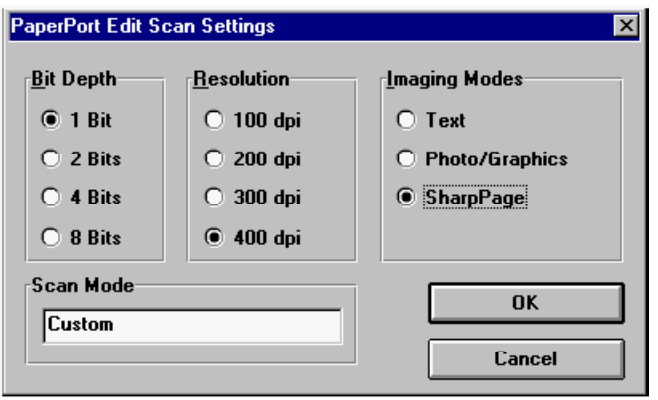
You are *not* recommended to use the scan modes "Faxing, Filing or Copying" (PaperPort “Deluxe”) or "Article" (PaperPort 3.5), which use the lower resolution of 200 dpi (and do *not* apply SharpPage image enhancement). Use this setting for “rough” black and white images which you want to archive, fax etc. Such images do not provide enough detail and quality to guarantee good OCR.

Occasionally, you may also use a **custom scan mode**. Custom settings are typically used to scan texts with a small point size (below 10 point) where a higher resolution may be needed.

Under PaperPort “Deluxe”, set the "Page Improvements" options - decide whether you want to deskew, trim and sharpen the image. You are recommended to automatically apply image deskewing with the option "Straighten".

Click the button "Advanced Settings" to determine the bit depth, resolution and imaging mode. It is imperative to leave the bit depth on 1 to generate black-and-white images: Readiris only accepts black and white images! Leave the imaging mode to "Text".

Under PaperPort 3.5, the same custom options are available, even if they are presented somewhat differently. When you select the scanning mode "Custom", the button "Show More" of PaperPort 3.5 is replaced by the "Edit Settings" button.



You are recommended to enable the imaging mode "SharpPage" rather than "Text": that image enhancement technology provides higher quality images.

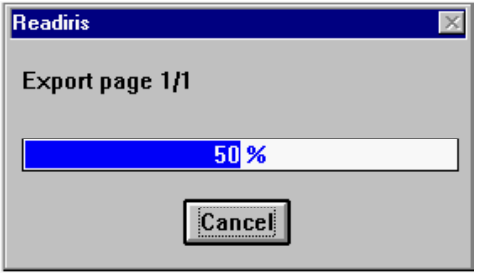
Custom scan modes are saved until they are modified again, even when you quit the PaperPort software.

Don't forget to set the **brightness** correctly! As PaperPort generates the images which are then passed on to a target application, it is here that all scan settings must be adjusted for optimal results.

Finally, don't forget to **calibrate** your scanner before using it!

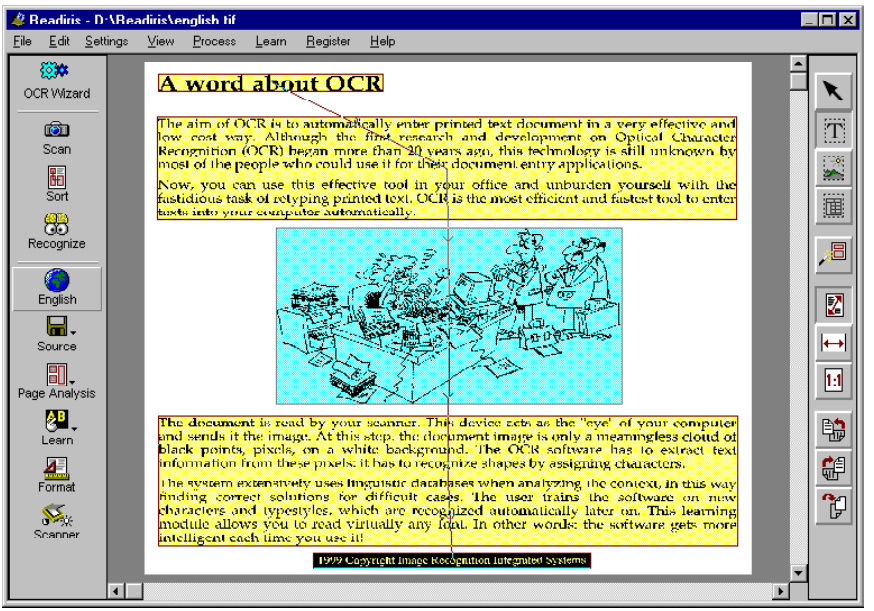
READIRIS OFFERS FULL FLEXIBILITY!

Let's try it now. Scan a page and drop an image thumbnail on the Readiris icon. When you do this, Readiris is started up and the selected image is exported.



We litterally mean image here, not images: Readiris only processes single pages and you are warned when you try to export several or stacked pages.

Page analysis is used to detect the various blocks - be they text, graphics or tables. As the system now stops, you can use all controls that Readiris offers to fine-tune the operation of the OCR software.



Thanks to Readiris **zoning** capabilities, you can limit the recognition to specific parts of a scanned page. You can for instance scan columnized pages from a magazine and OCR a single column. Where the page analysis has detected three text blocks, you can select only some for recognition with the "Sort" button. Or even erase all blocks and start from scratch by drawing windows around the text blocks of interest yourself.

Secondly, Readiris allows you to indicate the document's **language** - up to 55 languages are supported - and allows you to execute **font training** to increase the OCR accuracy even more - you can create font dictionaries to train the system on specific fonts, complete and activate existing font dictionaries to make good use of previous learning.

You can furthermore limit recognition to **numeric symbols** when reading tables of figures, indicate you want to read dot matrix printing etc.

Please refer to the Readiris manual to discover how fully featured and user controllable the Readiris software actually is!

There's one element the user *has* to take care of: saving the output. When you click the button "Recognize", the recognition is executed - with or without learning -, and the text result gets saved.

You can create text files in all common text formats, send the recognition result directly to an application - for instance Word - and copy the recognized text to the Windows clipboard. You can create an unformatted, running text and you can recreate the page **layout** of the original document.

When the recognition is completed, you return to the PaperPort desktop automatically.

However, if you dropped the scanned image on the wordprocessor icon - this can only be done with PaperPort 3.5 - or if you asked Readiris to send the recognition result to your wordprocessor, the recognized text isn't saved but sent on directly to the wordprocessor. If necessary, your wordprocessor is started up first.

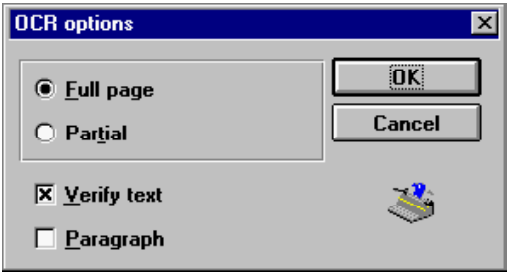
HOW ABOUT FULLY AUTOMATIC OCR?

This mode of operation offers full control over all settings. It is the default solution. However, you can also execute automatic OCR of full pages.

Whether you do or not depends on the OCR settings. This again functions somewhat differently under PaperPort 3.5 and the “Deluxe” versions.

With PaperPort “Deluxe”, things are very easy: disable the option "View after Scan" under the "Preferences" menu of Readiris and save the default settings with the command "Save Default Settings" under the "File" menu. As soon as you drop an image thumbnail on the Readiris icon, the image is exported and recognized promptly - page analysis is used to detect the text blocks.

With PaperPort 3.5, click the "Settings" button, select your wordprocessor in the "Category" list and click the button "OCR Settings". The OCR options are now displayed.



The setting "Full Page" or "Partial" indicates if OCR will be automatic or user controlled. Selecting "Full Page" leads to automatic OCR of entire pages: page analysis will detect and sort the text, graphics and table blocks to be recognized. The user may go through the learning phase but otherwise does not intervene. As soon as the OCR is executed, the wordprocessor is started up to display the text result.

The option "Partial" on the other hand leaves the user in control. That's the mode of operation which we illustrated previously.

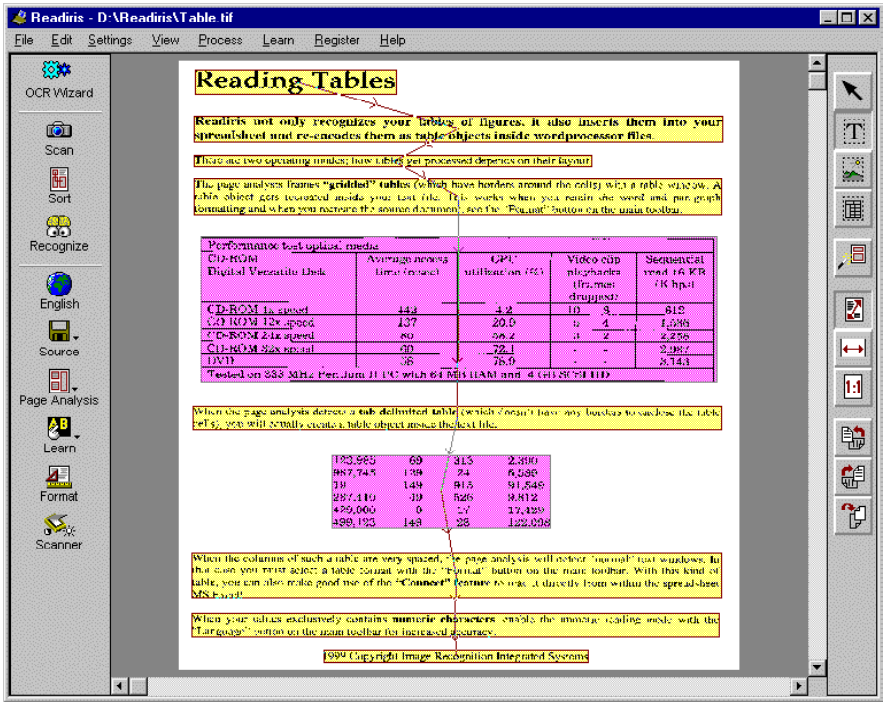
Two options are added: "Verify Text" and "Paragraph". Enabling the option "Verify Text" activates the learning mode, the option "Paragraph" ensures that no carriage returns are inserted at the end of each line, but only at the end of a paragraph. In other words, Readiris detects and recreates the text paragraphs, thus ensuring wordwrap for the recognized text. (It corresponds to the option "Merge Lines into Paragraphs" under the "Format" button of Readiris.)

For further details on font training and text formatting, refer to the Readiris User's Guide.

RECOGNIZING TABLES

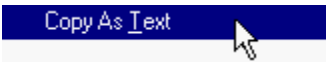
Readiris not only allows you to recognize texts but also tables of figures. With PaperPort "Deluxe", the table is dropped on the Readiris icon, with PaperPort 3.5, you can drop the scanned table onto the spreadsheet icon on the link bar to execute recognition.

The Readiris User's Guide explains how tables are detected by the page analysis, and how you can limit the recognition to numeric symbols.



PUTTING TEXT ON THE CLIPBOARD

We already mentioned that the Readiris link is not the only way of performing OCR under PaperPort 3.5. You can also select an item on your PaperPort desktop and select the command "Copy As Text" from the "Edit" menu.



The difference with the wordprocessor link is that the recognized text will not be transferred to your wordprocessor but copied to the clipboard. You can paste the result in your target application later on.

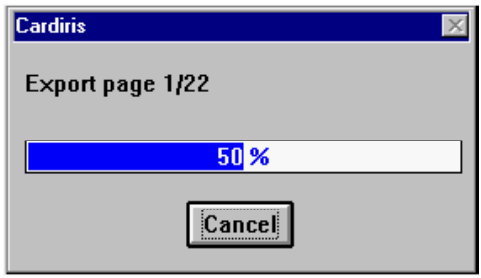
AND WHAT ABOUT CARDIRIS?

So much for the OCR software Readiris, on to the business card organizer Cardiris.

We proceed in the same way as with Readiris: PaperPort generates the images - the scan settings are taken care of by PaperPort - and “drag and drop” is used to send the images on to Cardiris. If necessary, Cardiris is started up automatically, the database last opened is opened again and the scanned images are exported to the Cardiris database.

When you have a **stack** of business cards to be archived, the best strategy is to scan all cards, staple them together with the "Stack" tool on the command bar and drop them all at once on the Cardiris icon.

The image transfer is indicated by a progress window.

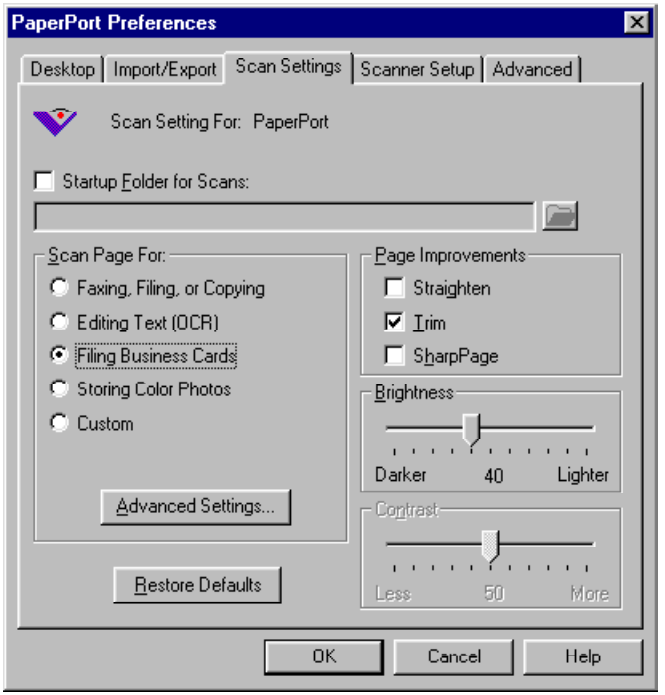


The PaperPort link limits itself to **archiving** the scanned business cards in Cardiris, it does *not* launch the recognition. However, the business cards are displayed within Cardiris so that the user can **index** them immediately. Don't forget that you can make good use of Cardiris' **batch OCR** capability to recognize a large number of cards with a single command!

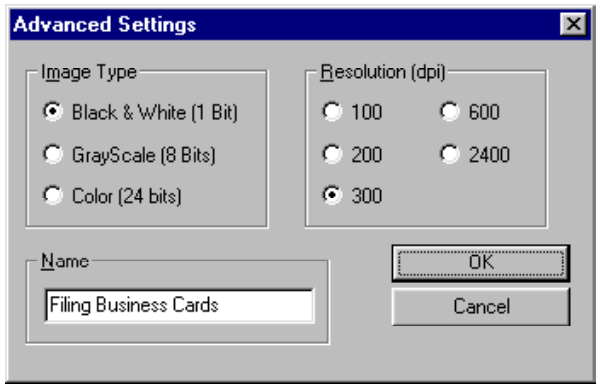
Efficiently archiving business cards is indeed just one of the many functions of Cardiris: text recognition allows to create a contact database automatically, you can search the stored cards intelligently, print lists and labels, export the contacts to other databases etc. Refer to the Cardiris User's Guide for full details.

PaperPort provides a predefined **scan mode** for business cards. As we already know, defining scan settings is done somewhat differently under PaperPort 3.5 and the “Deluxe” versions.

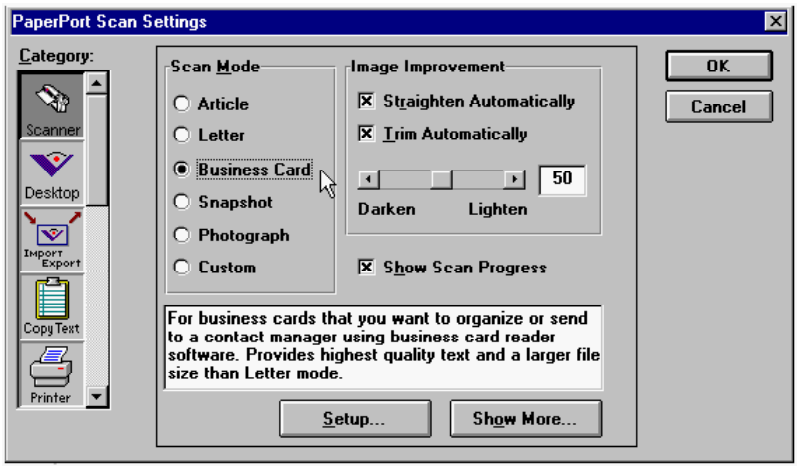
With the PaperPort “Deluxe” versions, click the "Settings" button on the command bar and select the scanning mode "Filing Business Cards".



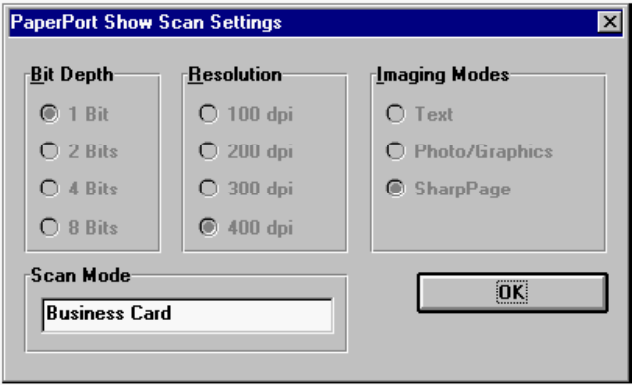
Clicking the button "Advanced Settings" reveals the details: with PaperPort 4.0, black and white images are scanned at a resolution of 400 dpi. This higher than usual resolution is useful given the tiny, stylized characters printed on business cards. (SharpPage image enhancement can be applied to generate optimized images.)



PaperPort 3.5 now. Click the "Settings" button on the command bar, select the scanner icon in the "Category" list and select the scan mode "Business Card". Enable the option "Trim Automatically" to limit the images to the actual business cards!



Click the button "Show More" to discover the details: black and white images are scanned at a resolution of 400 dpi - this higher than usual resolution is necessary given the tiny, stylized characters printed on business cards. SharpPage image enhancement is applied to generate optimized images.



IMPORTING PRESCANNED IMAGES

There's another way to OCR documents and organize business cards: importing prescanned images onto your PaperPort desktop.

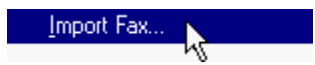
The command "Import" from the "File" menu allows to import **prescanned images** in various formats.

The original format has no relevance as the import converts them into PaperPort files anyway. The color mode on the other hand *is* relevant: don't forget that Readiris only accepts black-and-white images. You can export greyscale images to Cardiris, but these will be converted into black and white images!

(Other possibilities to import documents is printing them to the PaperPort desktop from their source applications and dragging an image file onto the PaperPort icon.)

IMPORTING AND RECOGNIZING FAXES

The command "Import Fax" from the "File" menu of PaperPort 3.5 allows you to import **faxes** received by the Symantec WinFax software. OCRing them with Readiris is certainly an option.



WHAT ABOUT ANNOTATIONS?

PaperPort items can be annotated: highlighting is applied, freehand or sticky notes are added, preprinted forms are marked up etc., all this is part of PaperPort's capabilities.

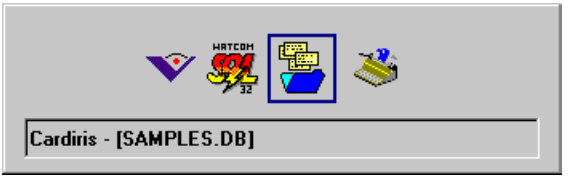
These annotations do not prevent recognition or business card archiving, but are not taken into account by Readiris and Cardiris. If a text was annotated, Readiris will recognize the original image *without* the annotations, when a business card was annotated, it is archived by Cardiris *without* the annotations.

CLOSING YOUR APPLICATIONS

PaperPort starts up your applications when needed, but doesn't close them again! Don't forget to close the applications which are no longer used.

To quickly access applications, you can always click on the button representing them on the taskbar.

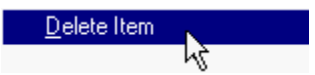
A much better way of proceeding is this: use the key combination Alt+Tab to browse through your open applications. This will allow you to close them one after the other quickly.



ERASING SUPERFLUOUS IMAGES

When you scan documents specifically to recognize them, the scanned images are no more than a temporary “vehicle”. As for Cardiris, that business card organizer archives calling cards very efficiently; the images can again be discarded from the PaperPort desktop.

Don't forget that all scanned images remain on the PaperPort desktop until the user decides to erase them. Unless you have good reasons to archive the images with PaperPort, you should delete superfluous images from your desktop with the command "Delete Item" under the "File" menu or the command "Delete" under the "Edit" menu to save disk space!



MAKING OPTIMAL USE OF PAPERPORT

This concludes our overview of the Readiris and Cardiris links for the PaperPort software.

Some additional information is given below on PaperPort functions which can be useful to the operation of Readiris and Cardiris. I.R.I.S. recommends you to make use of them given the right circumstances.

Remove Stray Dots (5.2) / Clean Page (lower versions)

The command "Remove Stray Dots" of PaperPort “Deluxe” 5.2 and the command "Clean Page" from the "Page" menu of earlier versions remove speckles, coffee stains, staple holes etc. that were picked up by the scanner. (This command is only available when the page view is enabled.)

Dirt smudges can decrease the speed and efficiency of the recognition process; speckles can get recognized as dots, comma's etc.

Speckles *outside* the text windows defined with Readiris have no relevance: zoning a scanned document correctly with Readiris renders page cleaning superfluous.

You can always clean a page manually as follows: activate the page view, choose the "Selection" tool on the annotation tool bar, select the zone to be removed and cut it from the page with the "Cut" command from the "Edit" menu.

Invert

The command "Invert" from the "Page" menu inverts the image: black “pixels” become white, white “pixels” become black. As a result, white text on a black background becomes legible. (This command is only available when the page view is enabled.)

Straighten (5.2) / Straighten Image (4.0) / Auto-Straighten Page (3.5)

The command "Straigthen" from the "Page" menu of PaperPort “Deluxe” 5.2, the command "Straigthen Image" from the "Page" menu of PaperPort “Deluxe” 4.0 and the command "Auto-Straighten Page" from the "Page" menu of PaperPort 3.5 straighten crooked pages by analyzing the image and rotating it with the best angle. (These commands are only available when the page view is enabled.)

Although there is some tolerance, exaggerated lineskew can slow down and even render impossible the recognition process.

This command is superfluous when the image improvement option "Straighten" (PaperPort “Deluxe”) or the option "Straighten Automatically" (PaperPort 3.5) is enabled in the scan settings.

To straighten a page manually, activate the page view, choose the "Straighten" tool on the annotation tool bar and draw a line to indicate at which angle you want to straighten the page.

Rotate Right, Flip and Rotate Left

The commands "Rotate Right", "Flip" and "Rotate Left" from the "Page" menu allow to modify the original orientation of the scanned documents.

Readiris and Cardiris are also equipped with rotation functions.

Some File Management

Let's end with the boring stuff: file management! The following files specifically concern the PaperPort link.

Readiris Folder

READPAP.GLK	PaperPort link
IRISOCR.OCR	OCR link

Cardiris Folder

CARDIRIS.GLK	PaperPort link

Cardiris Database Folder

MAX*.TIF	Image files

Card images (TIFF format) imported from PaperPort start with the character string *max*.

Windows Folder

MAXLINK.INI PaperPort initialization file

The PaperPort initialization file is updated by the Readiris and Cardiris installation program. It establishes the links with Readiris and Cardiris.