



Visioneer PaperPort

Readiris and Cardiris are trademarks of I.R.I.S.
OCR technology by I.R.I.S.

PaperPort is a trademark of Visioneer Communications Inc.
Hewlett-Packard is a trademark of Hewlett-Packard Co.

©1996 I.R.I.S. All rights reserved

Readiris, Cardiris and the Visioneer PaperPort software

This short manual discusses how to use the OCR package Readiris and business card organizer Cardiris with the Visioneer PaperPort scanning software. It contains all you need to know to get started with these I.R.I.S. products successfully using a Hewlett-Packard or Visioneer scanner.

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 and Cardiris software are installed using the **installation program** *setup.exe*.



Further details concerning their installation can be found in the Readiris and Cardiris manual.

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

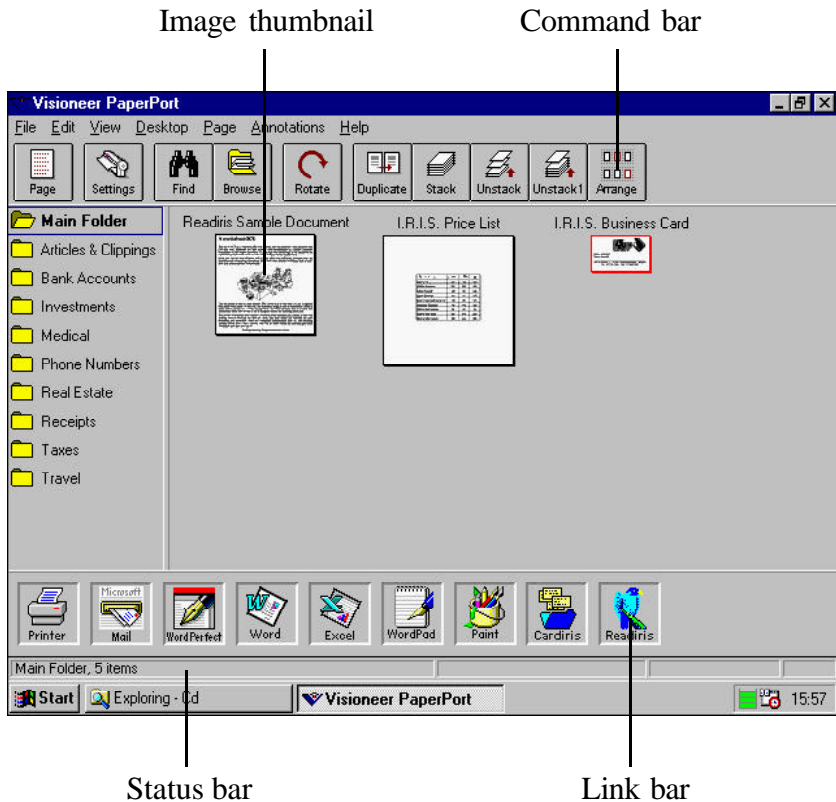
It doesn't even matter whether you install Readiris and Cardiris *before* or *after* you install the PaperPort software! Simply install these software packages on your PC and you can be up and running immediately.

STARTING UP THE PAPERPORT SCANNING SOFTWARE

Let's have a closer look at this. Start the PaperPort software by double-clicking the corresponding icon or by pushing the button on the left side of your sheetfed scanner.

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

USING PAPERPORT LINKS



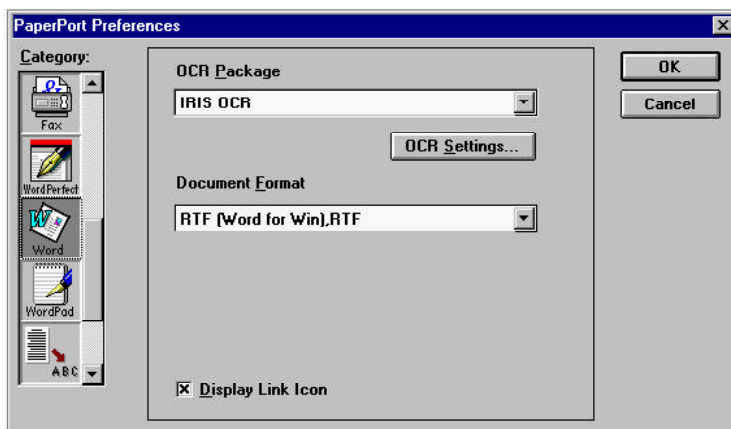
Actually, the Readiris OCR function is represented several times on the PaperPort desktop. Readiris is not only available as an icon on the link bar but is also “hidden” behind your wordprocessor and spreadsheet. As OCR converts images into editable text, your wordprocessor and spreadsheet take care of text recognition as well!

“Drag and drop” is how you do things with the scanned images: to OCR a *text*, it suffices to drag the image thumbnail onto the wordprocessor's icon. In this way, you indicate that you want to make use of the scanned document as a



text, and that character recognition has been executed. To OCR a *table* of figures, you drop its thumbnail on the spreadsheet icon, and 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.)

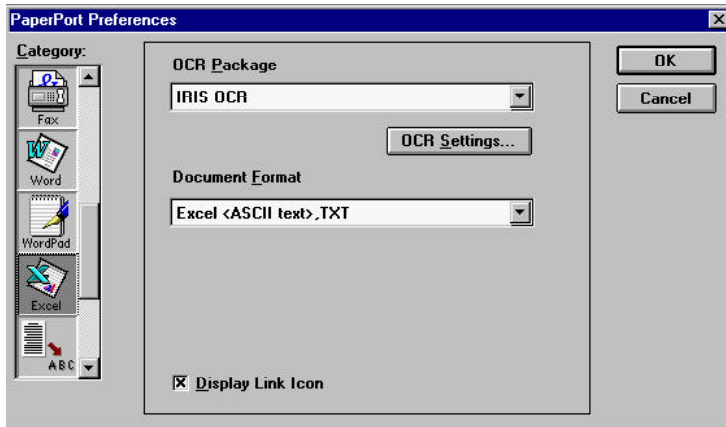
Click the "Settings" button on the command bar to see how Readiris and Cardiris were set up under PaperPort. Select your wordprocessor icon in the "Category" list now to see how Readiris is configured.



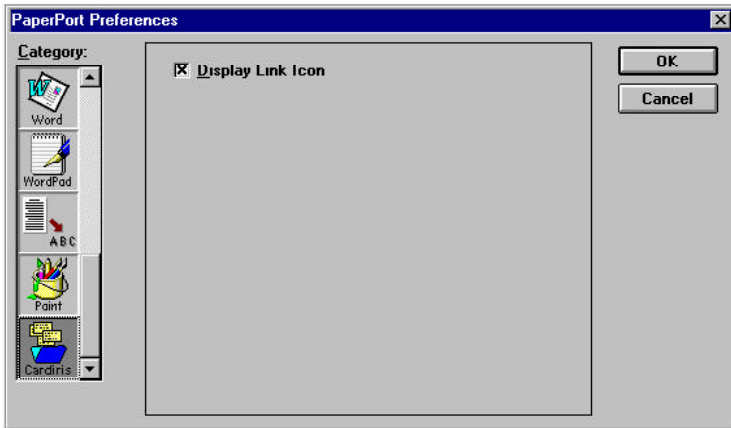
Readiris has been installed automatically as the OCR tool linked to your *wordprocessor* and the RTF text format will be used by default to transfer texts to your wordprocessor. You can select another file format, but this is again of no importance: the file format is only used internally by Readiris 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.

USING PAPERPORT LINKS



For *Cardiris* and *Readiris*, things are even simpler. When you select the Cardiris or Readiris icon in the "Category" list, a single option is displayed, "Display Link Icon". This option is enabled by default; leave it activated at all times, otherwise the Cardiris and Readiris application are no longer available on the link bar.



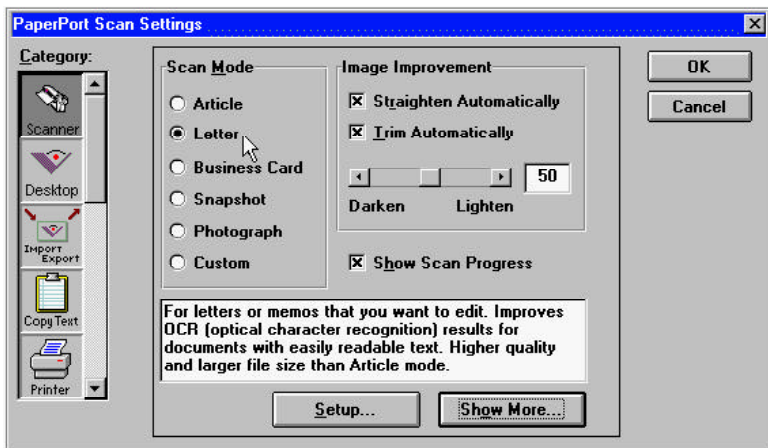
SETTING YOUR SCANNER FOR READIRIS

To OCR a document, scan it from the PaperPort desktop and drag its thumbnail onto the wordprocessor's icon. You can also drag a stack of images to OCR **multipage documents** with a single command.

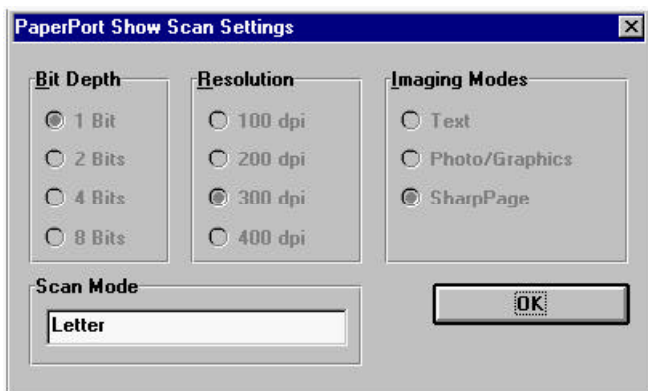
Be sure that you scan your document with the correct settings. Click the "Settings" button on the command bar to verify this. The scanning settings are displayed by default.

Select the **scan mode** "Letter". This is the standard setting for OCR purposes.

USING PAPERPORT LINKS



What these settings imply becomes clear when you click the button "Show More": black & white images are scanned at a resolution of 300 dpi with application of SharpPage image enhancement to optimize the images for the consecutive OCR process.

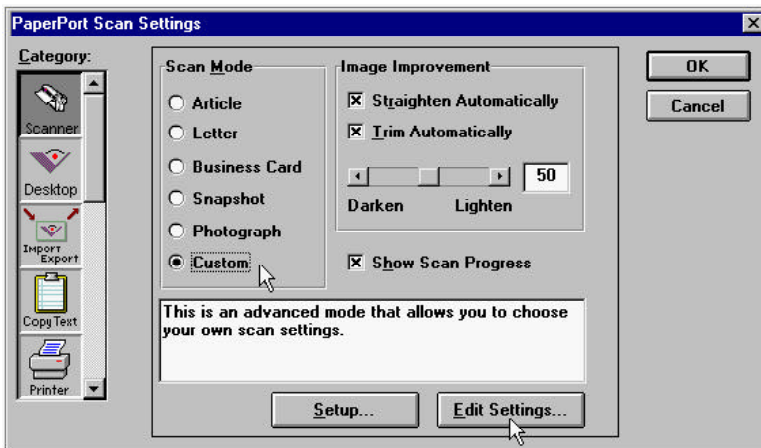




Note that the scan settings are always mentioned on the status bar.

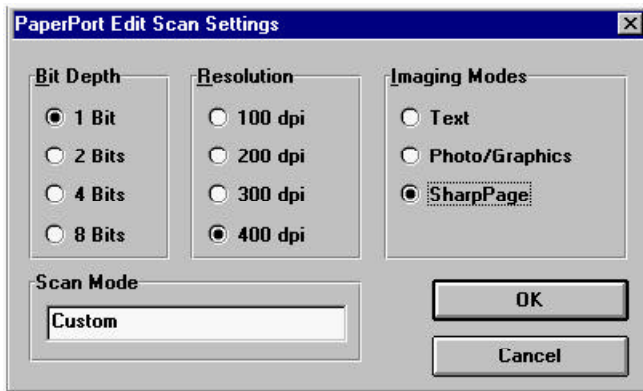
You are *not* recommended to use the scan mode "Article", which uses a lower resolution of 200 dpi and does not apply SharpPage image enhancement. Use this setting for “rough” black and white images which you want to archive, fax etc. Such images do not provide enough detail and quality to guarantee good OCR.

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



When you select the scanning mode "Custom", the button "Show More" is replaced by the "Edit Settings" button.

USING PAPERPORT LINKS

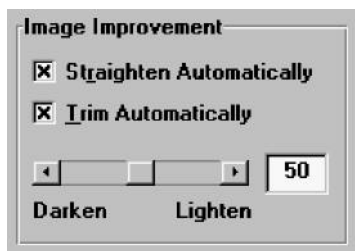


It is imperative to leave the bit depth on 1 to generate black and white images: Readiris only accepts black and white images. You are recommended to enable the imaging mode "SharpPage" rather than "Text": that image enhancement technology provides higher quality images.

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.

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



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

READIRIS OFFERS FULL FLEXIBILITY!

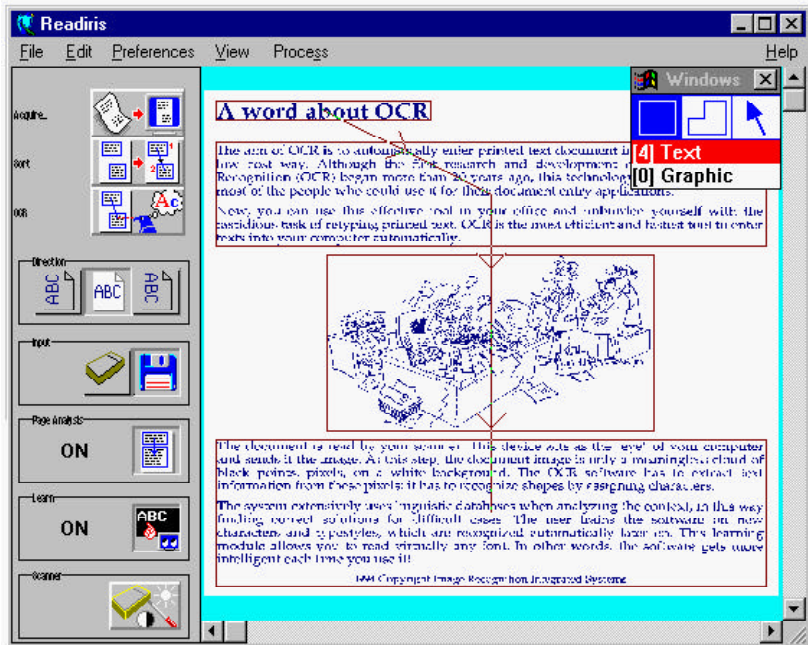
You are by no means limited to OCRing full pages as you scanned them. In fact, Readiris offers full control and high flexibility to the PaperPort user.

Thanks to Readiris **zoning** capabilities, you can limit the recognition to specific parts of a scanned page. You can e.g. scan columnized pages from a magazine and OCR a single column. And OCRing the first half of the left column, a single paragraph in the middle column and the lower half of the right column is accomplished with the same ease!

Furthermore, Readiris allows you to indicate the document's **language** - no less than 26 languages are supported - and allows you to execute **font training** to increase the OCR accuracy even more.

Let's see how all this works. When you drop an image thumbnail on the wordprocessor icon on the link bar, Readiris is started up automatically and displays the scanned image. Page Analysis is used to detect the various text blocks.

USING PAPERPORT LINKS



As the system now stops, you can use all controls that Readiris offers to fine-tune the operation of the OCR software. Where the Page Analysis has detected four 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.

You can also create user dictionaries to train the system on a specific font, complete and activate existing font dictionaries to make good use of previous learning, define the document's language or limit recognition to numeric symbols when reading tables of figures, indicate you want to read dot matrix printing, have the paragraphs and indentation detected and recreated etc.



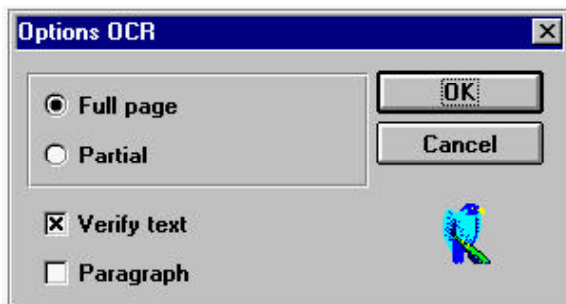
Please refer to the Readiris manual to discover how fully featured and user controllable the Readiris software actually is!

When you click the "OCR" button, the recognition is executed - with or without learning -, and the text result is sent on to the wordprocessor. If necessary, that 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 as defined under PaperPort. Click the button "Settings" on the command bar, 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 full pages: page analysis will detect and sort the text 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.

"Partial" on the other hand leaves the user in control. This is the mode of operation which we illustrated previously.

USING PAPERPORT LINKS

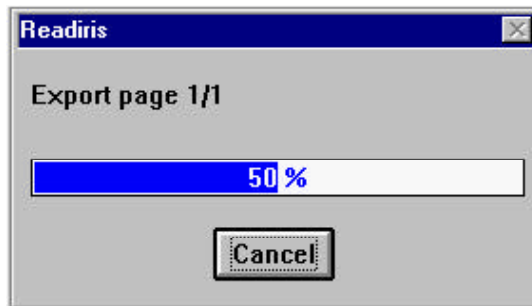
Two options are added: "Verify text" and "Paragraph". Enabling the option "Verify text" activates the learning mode, the option "Paragraph" ensures that no <Enter> codes are inserted at the end of each line, but only at the end of a paragraph. In other words, Readiris detects and recreates the text paragraphs, thus ensuring wordwrap for the recognized text.

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

ADDRESSING READIRIS DIRECTLY

The Readiris icon is also present on the link bar, so you can drop image thumbnails on the Readiris icon instead of on the wordprocessor or spreadsheet icon.

When you do this, Readiris is started up and the selected image is exported.



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



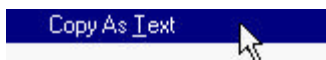
Readiris stops and gives you the opportunity of fine-tuning its operation. There's one element the user *has* to take care of: defining where he'll take the output. As the recognized text will not be sent automatically to the wordprocessor or spreadsheet, the user must determine this within Readiris.

You can create text files in all common text formats and copy the recognized text to the Windows clipboard. Refer to the Readiris User's Guide to learn how you are given full control over the output results.

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

PUTTING TEXT ON THE CLIPBOARD

The wordprocessor link is indeed not necessary to perform OCR. You can also select an item on your PaperPort desktop and select the command "Copy As Text" from the "Edit" menu.

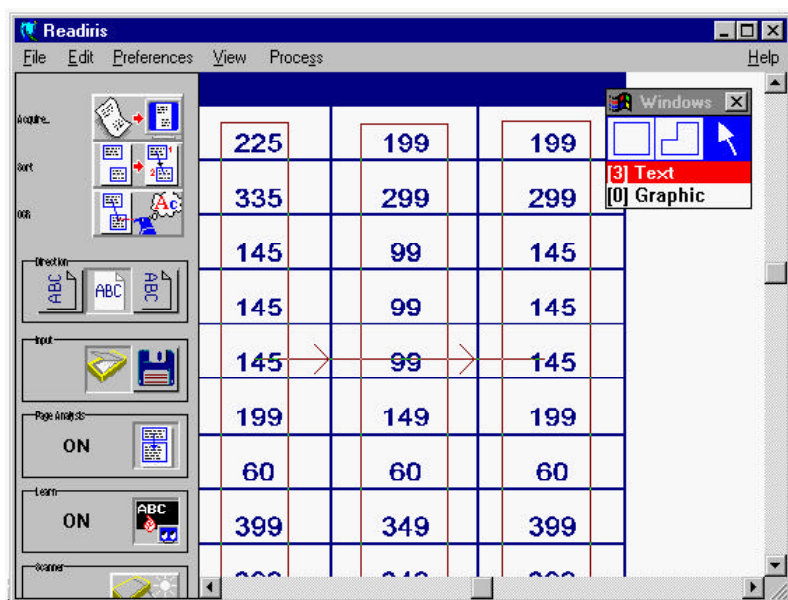


The difference with the wordprocessor link is that the recognized text will not be transferred to your wordprocessor but copied to the clipboard. You can paste the result in your target application later on.

RECOGNIZING TABLES

Readiris not only allows you to recognize texts but also tables of figures. Drop the scanned table onto the spreadsheet icon on the link bar to execute recognition.

The Readiris User's Guide tells you how to window tables correctly and how to limit the recognition to numeric symbols.





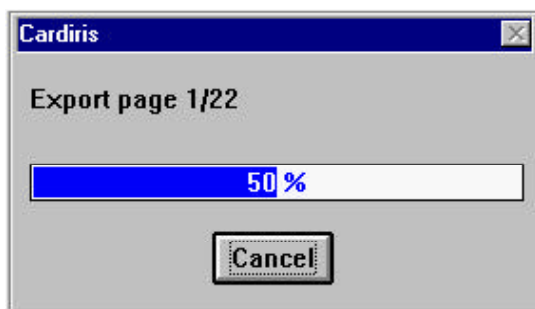
CARDIRIS AND PAPERPORT

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 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.



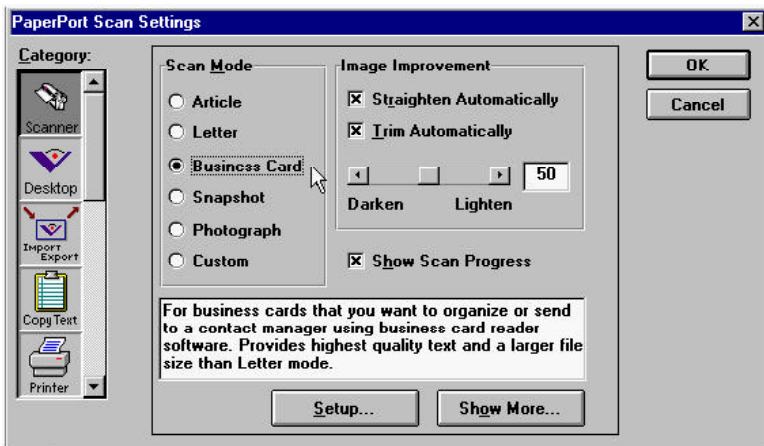
The PaperPort link limits itself to **archiving** the scanned business cards in Cardiris, it does *not* launch the recognition. However, the business cards are displayed within Cardiris so that the user can **index** them immediately. Don't forget that you can make good use of Cardiris' **batch OCR**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

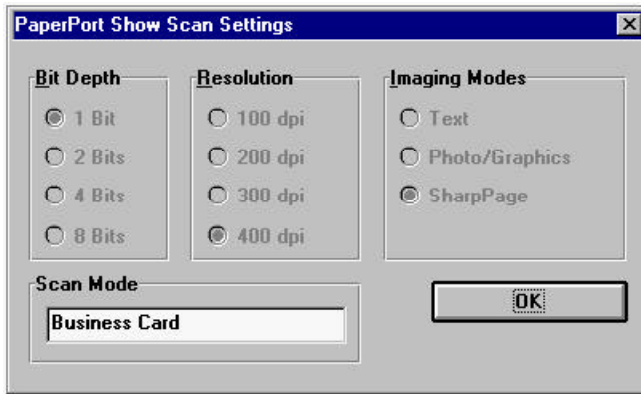
USING PAPERPORT LINKS

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. Click the "Settings" button on the command bar, select the scanner icon in the "Category" list and select the scan mode "Business Card".



Clicking the button "Show More" reveals 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 - and SharpPage image enhancement is applied to generate optimized images.



Be sure to enable the option "Trim Automatically" to limit the images to the actual business cards!

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.

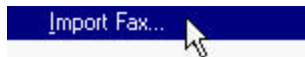


The original format has no relevance as the import converts them into PaperPort files anyway. The color mode on the other hand is relevant: don't forget that Readiris only accepts black and white images. You can export greyscale images to Cardiris, but these will be converted into black and white images!

(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 OCRING FAXES

The command "Import Fax" from the "File" menu allows you to import **faxes** received by the Delrina 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 (Windows 95) or minimize one application and restore another (Windows 3.1).

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 with the objective to OCR 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 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.

Clean Page

The command "Clean Page" from the "Page" menu removes speckles, coffee stains, staple holes etc. that were picked up by the scanner.

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: windowing a scanned document correctly with Readiris renders page cleaning superfluous.

To clean a page manually, 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.

Auto-Straighten Page

The command "Auto-Straighten Page" from the "Page" menu straightens crooked pages by analyzing the image and rotating it with the best angle.

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

This command is superfluous when the image improvement option "Straighten Automatically" 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.