ScanSoft®

OmniPage® SE

User's Guide

LEGAL NOTICES

IMPORTANT NOTICE

TRADEMARKS AND CREDITS

ScanSoft, Inc.
9 Centennial Drive
Peabody, MA 01960,
United States of America

C ONTENTS

Welcome

Welcome to OmniPage SE™, and thank you for using our software! The following documentation has been provided to help you get started and give you an overview of the program.

This User's Guide
This Guide introduces you to using OmniPage SE. It includes installation and setup instructions, a description of the program's commands and working areas, task-oriented instructions, ways to customize and control processing, and technical information. The Guide is presented in PDF format, allowing you to use hyperlink jumps on cross-references and other navigation tools in your PDF viewer.

Online Help
OmniPage SE's online Help contains information on features, settings, and procedures. The online Help is provided as HTML help, and has been designed for quick and easy information retrieval. Comprehensive context-sensitive help aims to provide just enough assistance to let you keep working without delay. Please see the section *Getting online help*.

Readme File
The Readme file contains last-minute information about the software. Please read it before using OmniPage SE. To open this HTML file, choose Readme in the OmniPage SE Installer or afterwards in the Help menu.

Scanning and other information
ScanSoft's web site at www.scansoft.com provides timely information on the program. The Scanner Guide contains up-dated information about supported scanners and related issues. Access ScanSoft's web site from the OmniPage SE Installer or afterwards from the Help menu.

USING THIS GUIDE

This Guide is written with the assumption that you know how to work in the Microsoft Windows environment. Please refer to your Windows documentation if you have questions about how to use dialog boxes, menu commands, scroll bars, drag and drop functionality, shortcut menus, and so on.

We also assume you are familiar with your scanner and its supporting software, and that the scanner is installed and working correctly before it is setup with OmniPage SE. Please refer to the scanner's own documentation as necessary.

The following conventions are used in this Guide:

Bold	Introduces new terms and presents sub-headings.
Italic	Names sections in this Guide (unless otherwise stated, the section is located in the same chapter as the reference). Names the main buttons used in automatic processing: *Start, Stop, Finish, Additional.*
Non-serif	Presents file names: sample.tif
Note	Presents an item of additional information.
Tip	Presents ideas for using program features to accomplish specific tasks.

GETTING ONLINE HELP
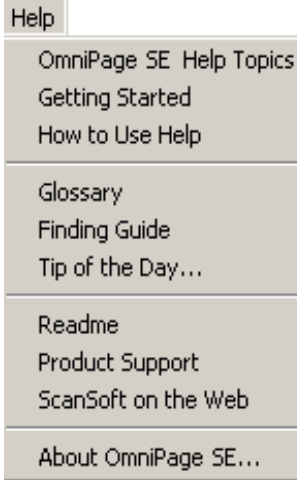
In addition to using this Guide, you can use OmniPage SE's online Help to learn about features, settings, and procedures. Online Help is available after you install OmniPage SE.

Online HTML Help

Open OmniPage SE's online Help at its top level by choosing OmniPage SE Help Topics at the top of the Help menu. This allows you to see topics arranged in a Table of Contents, search an alphabetical list of keywords or make full-text searches through the topics. Other items in the Help menu provide access to useful topics or web pages.

Press F1 as you are working with the program to see an online help topic relating to the current screen area, dialog box or warning message.

Context-Sensitive Help

You can get concise on-the-spot information in a popup window about a particular OmniPage SE menu item, toolbar button, screen area or dialog box, in the following ways:

Click the Help button in the Standard toolbar to get the help icon. Click this on any item on the desktop outside a dialog box or warning message.

Press Shift + F1 to get the same help icon.

Click the question mark button in the upper right corner of a dialog box and then click an item in the dialog box to see the popup window.
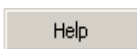
Some dialog boxes or warning messages have their own Help button, or a help text. Click the button or the text to get information on the dialog or message box.

Click anywhere to remove a context-sensitive popup Help window.

Tech Notes

ScanSoft's web site at www.scansoft.com contains Tech Notes on commonly reported issues using OmniPage SE. Web pages may also offer assistance on the installation process and troubleshooting.

Glossary

This Guide does not include a glossary. The online Help has a comprehensive glossary, with its own alphabetical index and a table of contents. Please consult it if you want to find the meaning of a term used in this Guide or in the program.

OMNIPAGE SE

The product you have is a Special Edition of the world-renown OmniPage Pro™ software. This edition has been developed for distribution by selected scanner manufacturers and contains a subset of the features of the OmniPage Pro 11 product. This Guide and the online Help describe the features of the full product, using an SE icon to document the differences between the two products.

If you find the additional features of the professional product would be of benefit to you, you can use online facilities to upgrade your Special Edition to OmniPage Pro 11.

1 Installation and setup

This chapter provides information on installing and starting OmniPage SE. It presents the following topics:

◆ System requirements

◆ Installing OmniPage SE

◆ Setting up your scanner with OmniPage SE

◆ How to start the program

◆ Registering your software

◆ New features in OmniPage Pro 11

◆ OmniPage SE and OmniPage Pro 11

SYSTEM REQUIREMENTS

You need the following minimum system requirements to install and run OmniPage SE:

◆ A computer with a Pentium or higher processor

◆ Microsoft Windows 95, Windows 98, Windows ME, Windows 2000, or Windows NT 4.0

◆ 32MB of memory (RAM), 64MB recommended

◆ 75MB of free hard disk space for the application files plus 10MB working space during installation

◆ 9MB for Microsoft Installer (MSI) if not present and 44MB for Internet Explorer if not present. (These are present as part of the operating system in Windows 98, Windows ME and Windows 2000.)

◆ SVGA monitor with 256 colors and 800 x 600 pixel resolution

◆ Windows-compatible pointing device

◆ CD-ROM drive for installation

◆ A compatible scanner if you plan to scan documents. Please see the Scanner Guide at ScanSoft's web site (www.scansoft.com) for a list of supported scanners.

Note Performance and speed will be enhanced if your computer's processor, memory, and available disk space exceed minimum requirements.

INSTALLING OMNIPAGE SE

OmniPage SE's installation program takes you through installation with instructions on every screen.

Before installing OmniPage SE:

◆ Make sure your scanner is connected, turned on, and compatible with your system.

◆ Close all other applications, especially anti-virus programs.

◆ Log into your computer with administrator privileges if you are installing on Windows 2000 or Windows NT.

◆ If you have previous OmniPage software on your system, the installer will ask for your consent to uninstall that software first.

▼ **To install OmniPage SE:**

1. Insert OmniPage SE's CD-ROM in the CD-ROM drive. The installation program should start automatically. If it does not start, locate your CD-ROM drive in Windows Explorer and double-click the Autorun.exe program at the top-level of the CD-ROM.

2. Choose a language to use during installation. This language will be used for the Text-to-Speech system and as the program's interface language. The program interface language is used for displays such as menu items, dialog boxes, warning messages and so on. You can change the interface language later from within OmniPage SE, but your choice at installation time determines which Text-to-Speech system will be installed with the program. References to the Text-to-Speech faciliy do not apply to OmniPage SE.

3. Follow the instructions on each screen to install the software. All files needed for scanning are copied automatically during installation.

Note Sometimes uninstalling and then reinstalling OmniPage SE will solve a problem. See *Uninstalling the software* at the end of chapter 6.

Note In OmniPage Pro 11, Text-to-Speech is available for English (British and US), French, German, Italian, Portuguese or Spanish. This is not available in OmniPage SE. See also the section *Reading text aloud* in chapter 4.

SETTING UP YOUR SCANNER WITH OMNIPAGE SE

All files needed for scanner setup and support are copied automatically during the program's installation. Before using OmniPage SE for scanning, your scanner should be correctly installed and tested for correct functionality.

Scanner installation and setup are done through the Scanner Wizard. You can start this yourself, as described below. Otherwise, the Scanner Wizard appears when you first attempt to perform scanning from OmniPage SE.

Please follow these steps to use the Scanner Wizard to setup your scanner with OmniPage SE:

◆ Choose Start➤Programs➤ScanSoft OmniPage SE➤ Scanner Wizard
or click the Setup button in the Scanner panel of the Options dialog box.
or choose a scan command in the Get Page drop-down list in the OmniPage Toolbox.

◆ Choose Select scanning source, then click Next.

◆ Click once on your scanner's TWAIN driver to select it, then click Next.

◆ Choose Yes to test your scanner configuration, then click Next.

◆ The wizard will now test the connection from the computer to your scanner. Click on Next.

◆ Insert a test page into your scanner.

◆ The wizard is now prepared to do a basic scan using your scanner manufacturer's software. Click on Next.

◆ Your scanner's native user-interface will appear. Click on Scan to begin the sample scan.

◆ If necessary, click on Inverse Image… or Missing Image… and make the appropriate selections.

◆ Once the image appears correctly in the window, click on Next.

◆ Select the item that most appropriately describes your scanner, then click on Next.

◆ Click on Next to proceed to page size.

◆ The page sizes that the Scanner Wizard believes that your scanner supports are listed in the window. To make any changes to the page sizes, click on Advanced, make the changes and then click on Next.

◆ Insert a page with text but no pictures into your scanner. Click on Next to begin a scan in black and white mode.

◆ If necessary, click on Inverse Image… or Missing Image… and make the appropriate selections.

◆ Once the image appears correctly in the window, click on Next.

◆ If you have a color scanner, insert a color photograph or a page with a color picture into your scanner. Click on Next to begin a scan in color mode. If necessary, click on Inverse Image… or Missing Image… and make the appropriate selections. Once the image appears correctly in the window, click on Next. If your scanner cannot scan in color, skip this step.

◆ Insert a photograph or a page containing a picture into your scanner. Click on Next to begin a scan in grayscale mode. If necessary, click on Inverse Image… or Missing Image… and make the appropriate selections. Once the image appears correctly in the window, click on Next.

◆ You have successfully configured your scanner to work with OmniPage SE! Click on Finish.

To change the scanner settings at a later time, or to set up a different scanner, or to test and repair an installed scanner, please follow one of these two methods to reopen the Scanner Wizard:

◆ Start▶Programs▶ScanSoft OmniPage SE▶Scanner Wizard or

◆ Start▶Programs▶ScanSoft OmniPage SE▶OmniPage SE▶Tools menu▶Options▶Scanner…▶Setup button.

Note To test and repair an improperly functioning scanner, follow the procedure above, selecting 'Test and configure current scanning source' at the start of the process.

HOW TO START THE PROGRAM

To start OmniPage SE do one of the following:

◆ Click Start in the Windows taskbar and choose Programs▶ScanSoft OmniPage SE▶OmniPage SE.

◆ Double-click the OmniPage SE icon in the program's installation folder or on the Windows desktop if you placed it there.

◆ Double-click an OmniPage Document (OPD) icon or file name; the clicked document is loaded into the program. See *OmniPage Documents* in chapter 2.

On opening, OmniPage SE's title screen is displayed and then its desktop. See chapter 2 for an introduction to OmniPage SE's desktop.

There are several ways of running the program with a limited interface:

◆ Use the Schedule OCR program. Click Start in the Windows taskbar and choose Programs▶ScanSoft OmniPage SE▶ Schedule OCR. See *Processing documents with Schedule OCR* in chapter 3.

◆ Click Acquire Text from the File menu of an application registered with the Direct OCR™ facility. See *How to set up Direct OCR* in chapter 3.

◆ Right-click an image file icon or file name for a shortcut menu. Select a sub-menu item from 'Convert To...' to define a target.

◆ Use OmniPage SE with ScanSoft's PaperPort® or Pagis® document management products, to add OCR services. See *How to use OmniPage SE with your PaperPort software* in chapter 3.

REGISTERING YOUR SOFTWARE

ScanSoft's registration Wizard runs at the end of installation. We provide an easy electronic form that can be completed in less than five minutes.

When the form is filled, and you click Send the program will search an Internet connection to immediately perform the registration online.

If you did not register the software during installation, you will be periodically invited to register later. You can go to www.scansoft.com to register online. Click on Support and from the main support screen choose Register on the left-hand column.

For a statement on the use of your registration data, please see ScanSoft's Privacy Policy.

NEW FEATURES IN OMNIPAGE PRO 11

The OmniPage® product family is augmented by OmniPage Pro 11 and OmniPage SE. This section lists enhancements introduced in the professional product OmniPage Pro 11. Some of these are incorporated in OmniPage SE, as detailed in the next section.

New features in OmniPage Pro 11 compared to OmniPage Pro 10 are:

◆ Greater accuracy - redeveloped recognition engines make OmniPage Pro 11 the most accurate OmniPage ever.

◆ Improved page layout - OmniPage Pro 11 will allow you to retain formatting that is true to the original, even on pages with non-gridded tables, headers and footers and dropped capitals.

◆ More intelligent proofreading - new IntelliTrain feature automatically uses previous corrections to generate better OCR results.

◆ PDF capability - now you can import PDF files (even read-only files) and convert them to your favorite program files (Word, Excel, etc.). You can also create PDF files from any paper document or image files.

◆ Better HTML - new WYSIWYG (What You See Is What You Get) HTML output will handle graphics, text, and backgrounds to keep your web output looking like the original document.

◆ Language support - OmniPage Pro 11 now supports over 100 languages and extends to the Greek and Cyrillic alphabets.

◆ Detail view - this provides more customizable information about each page, making it easier to handle pages in a document.

◆ Text Editor - a new fully-featured WYSIWYG editor for recognition results, with a wide range of editing tools, color support, and a choice of four formatting levels for display and export.

◆ Better results on degraded text - a new despeckle module significantly reduces errors on spotty, shaded and color backgrounds.

OMNIPAGE SE AND OMNIPAGE PRO 11

This list documents features which are not incorporated in OmniPage SE, but which can become available by upgrading to OmniPage Pro 11:

◆ Significant improvement in recognition accuracy.

◆ Access to the IntelliTrain character training facility.

◆ Abitity to open and read the contents of PDF files.

◆ Ability to save recognized documents to PDF format.

◆ Ability to open TIFF FX image files.

◆ Handling LZW TIFF and GIF image files for input and output.

◆ Support for WYSIWYG HTML 4.0 output.

◆ Language support rises from about 50 to over a hundred.

◆ Access to text-to-speech software, allowing recognized texts to be read aloud.

For more information or to upgrade, please visit www.scansoft.com, make a selection from the country/continent list if you prefer a different language, then click on the OmniPage icon.

2 Introduction

You probably use your computer for business correspondence, preparing reports, handling data and an ever-increasing number of other uses. The challenge is that, in spite of the digital revolution, certain sources of information still circulate in printed, paper form and cannot be used immediately in a computer.

For example, if you want to incorporate information from a magazine article in a report you are preparing, you somehow have to get the text from the article into your computer. Painstakingly retyping the article is not an appealing solution.

This chapter introduces you to the solution: optical character recognition (OCR). It describes how OmniPage SE uses OCR technology to transform text from scanned pages or image files into editable text for use in your favorite computer applications.

The chapter includes the following sections:

- ◆ What is optical character recognition
  - ◆ Documents in OmniPage SE
  - ◆ Basic processing steps
- ◆ The OmniPage SE desktop
- ◆ Managing documents
- ◆ OmniPage Documents
- ◆ Settings

WHAT IS OPTICAL CHARACTER RECOGNITION
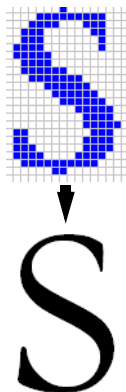
Optical character recognition is the process of extracting text from an image. This image can result from scanning a paper document or opening an electronic image file. Images do not have editable text characters; they have many tiny dots (pixels) that together form character shapes. These present a picture of the text on a page.

During OCR, OmniPage SE 11 analyzes the character shapes in an image and defines solutions to produce editable text. After OCR, you can save the resulting text to a variety of word-processing, desktop publishing or spreadsheet applications.

OmniPage SE's OCR capabilities

In addition to text recognition, OmniPage SE can retain the following elements of a document through the OCR process.

Graphics
Photos, logos, and drawings are examples of graphics.

Text formatting
Font types, sizes and styles (such as **bold**, *italic* and underlines) are examples of character formatting. Indents, tabs, margins and line spacing are examples of paragraph formatting.

Page formatting
Column structure, table formats, and placement of graphics and headings are examples of page formatting.

The graphics, text and page formatting elements that OmniPage SE retains are determined by the settings you select. Refer to the *Settings Guidelines* in the online Help for more information about selecting settings.

Note OmniPage SE only recognizes machine-generated characters such as offset or laser-printed or typewritten text. However, it can retain handwritten text, such as a signature, as a graphic.

Documents in OmniPage SE

OmniPage SE handles documents one at a time. When you acquire your first image (from scanner or from file) a new document is started. Further acquired images are added to the same document, until you save and close it.

A document in OmniPage SE consists of one image for each document page. After you perform OCR, the document will also contain recognized text, displayed in the Text Editor, possibly along with graphics and tables. For more information on screen areas, see the section *The OmniPage SE desktop.*

Basic processing steps

There are two main ways of handling documents: with automatic processing or manual processing. See chapter 3, *Processing documents automatically* and *Processing documents manually.* The basic steps for both processing methods are broadly the same:

1. **Bring a set of images into OmniPage SE.**
   You can scan a paper document with or without an Automatic Document Feeder (ADF) or load one or more image files. The resulting images appear in miniature in the Document Manager's Thumbnail view and the pages are summarized in its Detail view. The image of the current page is displayed in the Original Image area.

2. **Perform OCR to generate editable text.**
   During OCR, OmniPage SE creates **zones** around elements on the page that will be processed, and then interprets text characters or graphics in each zone. Manual and template zoning are also possible. After OCR, you can check and correct errors in the document using the OCR Proofreader and edit the document in the Text Editor.

3. **Export the document to the desired location.**
   You can save your document to a specified file name and type, place it on the Clipboard, or send it as a mail attachment. You can save it as an OmniPage Document (OPD) as described later. You can save the same document repeatedly to different destinations, different file types, with different settings and levels of formatting. See chapter 5.

THE OMNIPAGE SE DESKTOP

OmniPage SE's desktop has a title bar and a menu bar along the top and a status bar along the bottom. It has three main working areas, separated by splitters: the Document Manager, the Original Image area and the Text Editor. The Document Manager has two tabbed panels: Thumbnail view and Detail view. The Original Image area has an Image toolbar and the Text Editor has a Formatting toolbar.

Formatting toolbar

Standard toolbar

OmniPage Toolbox

The current page has a pale border.

This page has been recognized.

Thumbnail view shows a picture of each page in the document.

Page navigation buttons

Buttons to show, hide or rearrange the working areas.

Image toolbar

Drag this splitter to left or right to resize the working areas.

The Text Editor view buttons offer four formatting levels.

Original Image area:
This displays the image of the current page, together with any zones automatically or manually placed on the image.

Text Editor: This is displaying the recognition results from the current page in True Page™ view.
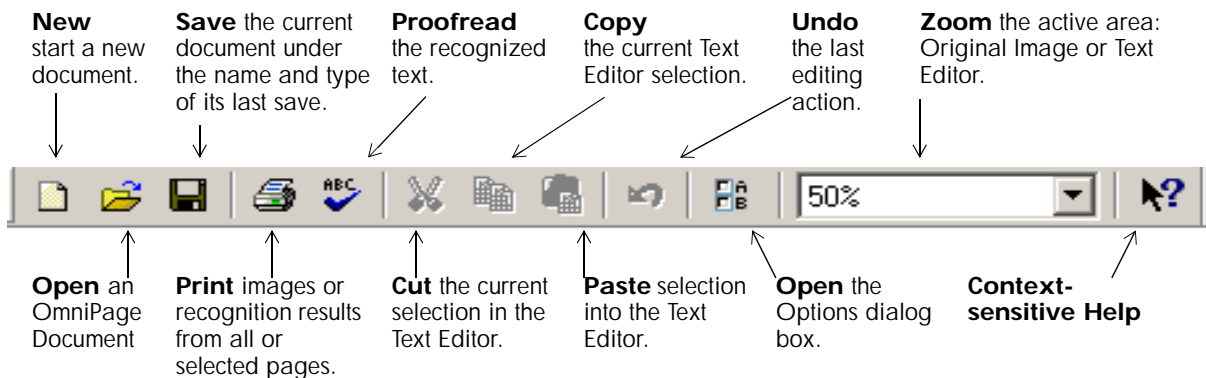
Note   To control which of the three views (Document Manager, Original Image, and Text Editor) are displayed, check or uncheck each view from the View menu or with the status bar buttons.

The OmniPage Toolbox lets you control processing. It can have three states, depending which of the three tab buttons on the left is clicked. In the picture, we display its appearance for Manual OCR. We show the program with a three-page document. Page one is the current page, which has been recognized and proofed. Page two has been recognized but not proofed yet. Page three has been acquired and manually zoned, but not recognized yet. The icons at the bottom right of the thumbnail images show page status.

Status bar buttons let you show, hide or rearrange the main screen areas and move to other pages in the document. A right mouse click in any screen area brings up a shortcut menu with the most useful commands for that area.

The Standard toolbar

The Standard toolbar contains buttons and a drop-down list for performing standard tasks. It can be floated and docked to any edge of the OmniPage SE desktop. All these functions can also be accessed from menus.

New start a new document.

Save the current document under the name and type of its last save.

Proofread the recognized text.

Copy the current Text Editor selection.

Undo the last editing action.

Zoom the active area: Original Image or Text Editor.

Open an OmniPage Document

Print images or recognition results from all or selected pages.

Cut the current selection in the Text Editor.

Paste selection into the Text Editor.

Open the Options dialog box.
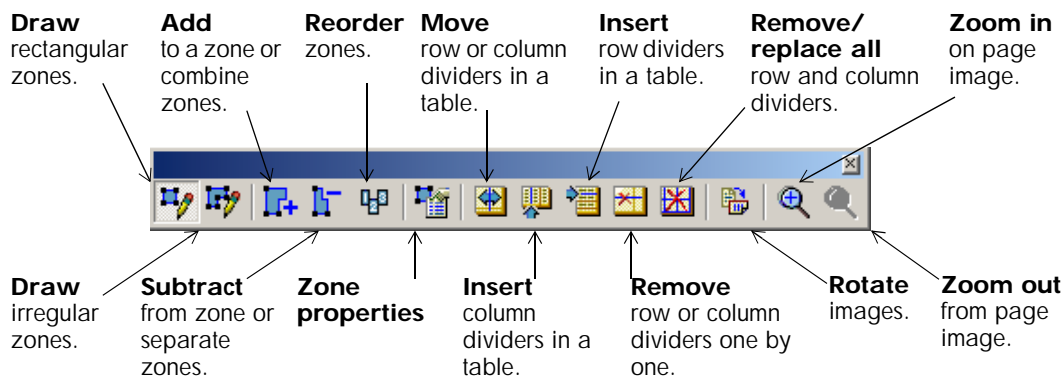
Context-sensitive Help

The Menu bar

For concise information on any menu item, click the context-sensitive help button and then click a menu item. A popup text explains the purpose of the menu item. Click anywhere to close the popup.
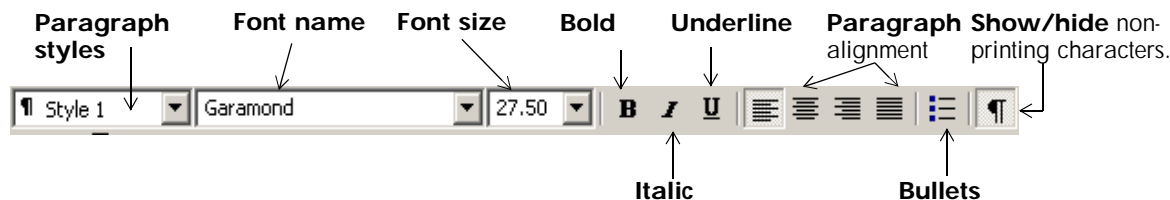
The Image toolbar

The Image toolbar contains buttons that allow you to zoom in or out on the current image or to rotate it. They also allow you work with zones and table dividers on the page. See chapter 3, *Manual zoning* and *Table grids in the image*. Here we summarize the purpose of the buttons. The Image toolbar can be floated (that is, undocked and moved anywhere on the desktop). It can be docked to any edge of the Original Image area.

Draw rectangular zones.

Add to a zone or combine zones.

Reorder zones.

Move row or column dividers in a table.

Insert row dividers in a table.

Remove/ replace all row and column dividers.

Zoom in on page image.

Draw irregular zones.

Subtract from zone or separate zones.

Zone properties

Insert column dividers in a table.

Remove row or column dividers one by one.

Rotate images.

Zoom out from page image.

Tip You can also resize or rotate the original image with a shortcut menu. Right click in the Original Image area outside a zone and select a zoom or rotation value.

The Formatting toolbar

The Formatting toolbar contains buttons that allow you to edit recognized text in the Text Editor. See *Text and image editing* in chapter 4. Here we summarize the purpose of the buttons. The Formatting toolbar always remains along the top of the Text Editor.

Paragraph styles

Font name

Font size

Bold

Underline

Paragraph alignment

Show/hide non-printing characters.

¶ Style 1 | Garamond | 27.50 | **B** *I* U | ≡ ≡ ≡ ≡ | ≔ | ¶

Italic

Bullets

The OmniPage Toolbox

This Toolbox lets you drive the processing. By default it is located along the top of the OmniPage SE desktop, just above the working areas. It can be floated and also be docked along the bottom of the desktop.

It has three tabs on the left: AutoOCR™, Manual OCR and OCR Wizard. Click one to see its controls in the Toolbox. The picture at the beginning of this section showed the OmniPage desktop with the Manual OCR toolbar. The AutoOCR toolbar looks like this.

Automatic processing is started, and can be stopped and re-started with the buttons on the right of the toolbar. The use of these buttons is explained in *Processing documents automatically* in chapter 3. The effects of other settings are also described in chapter 3, *Tutorial: Processing documents*

You can switch between automatic and manual processing any time the program is not busy with processing. That means you can switch between them while you are working within a document. You can automatically process some pages, then add more pages with manual processing. After processing a stack of pages automatically, you can inspect the results and then go back to reprocess certain pages manually. This procedure is described in chapter 3 in the section *Processing a document automatically and finishing it manually*.

OmniPage SE must be empty when you start the OCR Wizard. See the section *Processing documents using the OCR Wizard* in chapter 3. When you have used the OCR Wizard to process and save a document, it remains in the program and can be further processed (adding more pages, rerecognizing pages etc.) with either manual or automatic processing.

MANAGING DOCUMENTS

The Document Manager is situated on the left of the OmniPage SE desktop. It has two tabbed panels: Thumbnail view and Detail view. Click a tab to see its view. Both views summarize the pages in the document and are synchronized: the current and selected pages remain the same when you switch views. Our pictures show the two views with the same four-page document. Pages 1 and 2 are selected and page 4 is the current page, that is, the one shown in the Original Image area. The Document Manager shows page status with the following icons:

Page	Status	Thumb-nail icon	Detail icon	Page image has been...
1	Acquired	—	📄	Acquired with no manual or template zones and has not yet been recognized.
2	Zoned	⬜	—	Acquired and manual or template zones have been placed; not yet recognized.
3	Recognized	👓	🔍	Recognized, but not proofread, or proofing was interrupted on the page.
4	Proofed	👓✓	🔍✓	Recognized, and proofing has reached the end of the page.
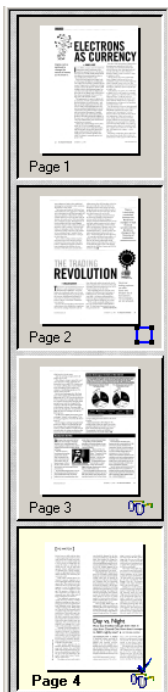
Thumbnail view

This presents a vertical set of numbered thumbnail images, one for each page in the document. Scroll to see pages as necessary. The current page has a paler background and its page number text appears bold. You can select multiple pages in the document; these have a 'pushed-in' appearance. A status icon appears at the bottom right of each page as described above.

Jump to a page: Click the icon of the desired page.

Reorder a page: Click the thumbnail of the page you want to move and drag it above the desired page number. Pages are renumbered automatically.

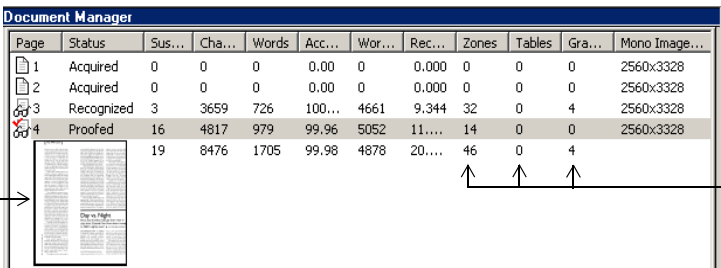**Delete a page:** Select the thumbnail of the page you want to delete and press the Delete key.

Select multiple pages: Hold down the Shift key and click two thumbnails to select all pages between and including them. Hold down the Ctrl key as you click thumbnails to add pages to a selection one by one. Then you can move or delete the selected pages as a group, or send them to (re)recognition.

Detail view

This facility is new to OmniPage SE. It provides an overview of your document with a table. Each row represents one page. Columns present statistical or status information for each page, and (where appropriate) document totals. The picture below shows the default columns on the left and four columns which a user has specified.

Move the cursor onto the page's status icon to see a thumbnail of the page.

Document Manager											
Page	Status	Sus…	Cha…	Words	Acc…	Wor…	Rec…	Zones	Tables	Gra…	Mono Image…
1	Acquired	0	0	0	0.00	0	0.000	0	0	0	2560x3328
2	Acquired	0	0	0	0.00	0	0.000	0	0	0	2560x3328
3	Recognized	3	3659	726	100…	4661	9.344	32	0	4	2560x3328
4	Proofed	16	4817	979	99.96	5052	11.…	14	0	0	2560x3328
		19	8476	1705	99.98	4878	20.…	46	0	4	

This shows the number of zones of each type on the page.

The current page is shown with a highlight. You can use Detail view for page operations, as follows:

Jump to a page: Click the row of the desired page.

Reorder a page: Click the row of the page you want to move and drag it to the desired location. An arrow indicator on the left shows where the page will be inserted. Pages are renumbered automatically.

Delete a page: Select the row of the page you want to delete and press the Delete key.
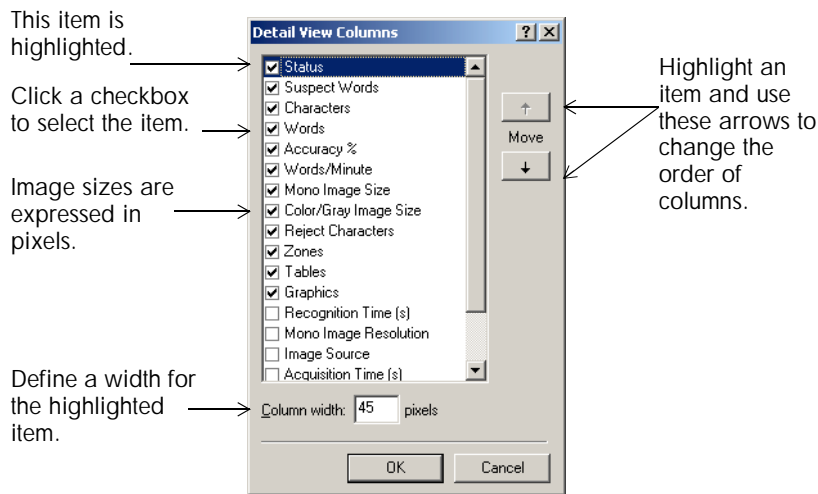
Select multiple pages: Hold down the Shift key and click two page rows to select all pages between and including them. Hold down the Ctrl key as you click rows to add pages to a selection one by one. Then you can move or delete the selected pages as a group, or send them to (re)recognition.

When multiple pages are being selected, the page set as current does not change. All selected pages are highlighted.

Tip Get image size information by hovering the cursor over a thumbnail or outside a zone on an original image. A popup text displays the image size in pixels and the program's unit of measurement. Image resolution is also shown.

Customizing columns in Detail view

You can specify which columns of information you want to see in Detail view. Click Customize Details... in the View menu for the following dialog box:

This item is highlighted.

Click a checkbox to select the item.

Image sizes are expressed in pixels.

Define a width for the highlighted item.

Highlight an item and use these arrows to change the order of columns.

Detail View Columns

- ☑ Status
- ☑ Suspect Words
- ☑ Characters
- ☑ Words
- ☑ Accuracy %
- ☑ Words/Minute
- ☑ Mono Image Size
- ☑ Color/Gray Image Size
- ☑ Reject Characters
- ☑ Zones
- ☑ Tables
- ☑ Graphics
- ☐ Recognition Time (s)
- ☐ Mono Image Resolution
- ☐ Image Source
- ☐ Acquisition Time (s)

Move ↑ ↓

Column width: 45 pixels

OK   Cancel

Define which columns should appear, their widths, and column order. The topic *Customizing Detail view columns* in online Help clarifies what is presented in each column. You can change column widths easily in Detail view; just drag the column dividers in the title bar.

Deleting pages from a document

Page deletions must be confirmed and can be undone. Delete the current page only with the item Delete Current Page in the Edit menu. Delete all selected pages in the Document Manager (either view) by pressing the Delete key or using the shortcut menu command Clear.

Printing a document

You can print the document with the Print item in the File menu. Choose whether to print images or text (that is, recognition results as they appear in the Text Editor). You can print all pages or a range of pages. The Print button in the Standard toolbar prints images or text, depending whether the Original Image area or the Text Editor is active.

Closing a document

Choose Close in the File menu to close a document. You are prompted to save your document if you have not saved it or you have modified it since the last save. See the next section on saving the document as an OmniPage Document (*.opd). You will also be prompted to save unsaved training data if you selected 'Prompt to save IntelliTrain' data when closing document' in the Proofing panel of the Options dialog box.

The last sentence does not apply to OmniPage SE.

OMNIPAGE DOCUMENTS

The OmniPage Document is the program's proprietary file type; it has the extension .opd. It is one of the file types offered when saving a document to file. You save the document to the OPD file type if you want to work with it again in OmniPage SE during a future session. You can then process unfinished pages, add more pages and proof or edit recognition results.

An OmniPage Document contains the original page images with any zones placed on them. After recognition, the OPD also contains the recognition results. Recognized characters are stored along with their coordinate and confidence data. This preserves the links between image and text, so that verification and proofing remain available when the OPD is reopened in future sessions.

When you save an OmniPage Document, the current settings (and unsaved training) are also saved. When you open an OmniPage Document, its settings are applied, temporarily replacing those existing in the program.

Why save to OPD

You do not have to save your documents to the OPD file type. You would typically do this for the following reasons:

♦ You cannot finish working with the document in the current session.
♦ You want to pass the document to other users who have OmniPage SE or OmniPage Pro 11. For example, you can pass an OPD file to a specialist for proofing. In an office network, you may have one scanner generating images for recognition and proofing at several workstations.
♦ You want to build up an archive of recognized documents whose original images remain accessible. The recognized texts allow searching by keywords and other document retrieval techniques.

Note Recognition results should be saved away from OPD files before installing any OmniPage upgrade. These files may not be upwards compatible to newer OPD file formats, or possibly only the images will be retained when the files are upgraded.

How to save to OPD

If you intend to create an OPD, you can save it to this format at an early stage, for protection. Use the Save button to save it periodically as you work. Save it again at the end of your session.

The Save button saves the document to the name and file type of its last save. You can save your document repeatedly to different formats. If your first save was to another format (for instance .DOC), use the item Save As... from the File menu to save it as an OPD. If a document is saved as an OPD, then you later save it to another format, it is not automatically resaved as an OPD. When you close the document or exit the program, you will be prompted to save the document as an OPD.

SETTINGS

The Options dialog box is the central location for OmniPage SE settings. It has seven panels. Context-sensitive help provides information on each setting. In overview, the settings panels are:

OCR
Use this to specify recognition language(s), a user dictionary, a reject character, an OCR method (optimize for speed or accuracy) and font matching.

Scanner
Use this to define page size and orientation for scanning. You can also make brightness and contrast settings and define options for scanning multi-page documents, with or without an Automatic Document Feeder (ADF). You can change scanner setup settings or install a new scanner or change the default scanner.
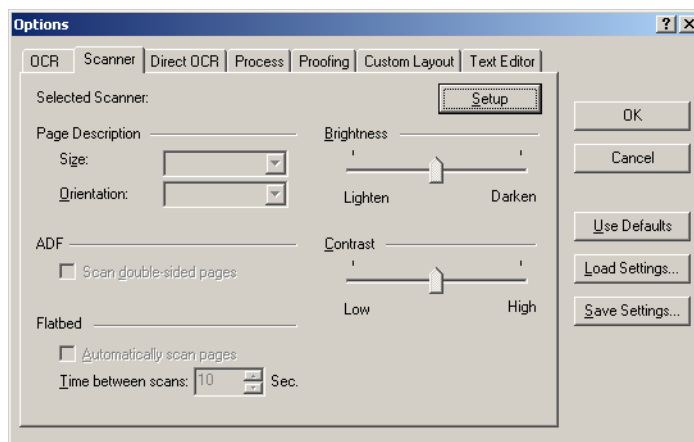
Direct OCR™
This feature provides OCR services directly from your favorite word processor or similar application. Use this panel to register and unregister applications for Direct OCR and to enable or disable this service. You can also specify automatic or manual zoning and whether proofreading is desired or not.

Process

Use this to define where new images should be placed in the document and set other preferences governing the behavior of the processing. You can change the interface language here.

Proofing

Use this to define whether proofreading should begin automatically after recognition. Define also whether IntelliTrain should run, and use it to load or work with a training file. For more detail, see chapter 4, *Proofreading OCR results.*

The references to IntelliTrain and training files do not apply to OmniPage SE.

Custom Layout

Use this to describe the layout of your input document pages very precisely. This gives you maximum control over the auto-zoning process, instructing it to search or ignore columns, graphics and tables.

Text Editor

Use this to show or hide some features in the Text Editor, to define the unit of measurement to be used and to turn word wrapping on or off.

Note   Some settings have an effect only on future recognition. Examples are the recognition languages, a training file and scanner brightness. These settings should be correctly adjusted before you start processing. To have changes in these settings applied to already recognized pages, you will have to rerecognize them. Other settings are implemented immediately in all existing pages. Examples are Text Editor settings like word wrap and measurement units.

---

3 Tutorial: Processing documents

This chapter describes different ways you can process a document and also provides information on key parts of this processing.

- ◆ Quick Start Guide
- ◆ Processing documents using the OCR Wizard
- ◆ Processing documents automatically
- ◆ Processing documents manually
- ◆ Processing a document automatically and finishing it manually
- ◆ Processing from other applications
- ◆ Processing documents with Schedule OCR

The detailed topics are:

- ◆ Defining the source of page images
- ◆ Describing the layout of the document
- ◆ Manual zoning
- ◆ Table grids in the image
- ◆ Using zone templates

QUICK START GUIDE

This topic takes you step-by-step through the basic OCR process.

Loading and recognizing sample image files

You will find sample image files in the program folder, both single-page and multi-page files. First try reading these files using the procedure presented below, except for the references to a scanner. See *Input from image files* for more information on acquiring the images. The results provide you with a benchmark of the recognition quality you should expect from your own files of comparable quality.

Next, try scanning a page from your scanner.
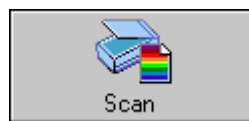
Scanning and recognizing a single page

Turn your scanner on and be sure it is working correctly. Choose a page with good-quality clear text for this test.

We assume OmniPage SE's default settings are set and that your document is in the language you specified for interface language during installation. Open the Options dialog box from the Tools menu and choose Use Defaults if you are not using the program for the first time.

You will process the document automatically and save the recognition results to a file. You will proof the document but will not edit it inside OmniPage SE's Text Editor.

	What you do	What happens
1.	Set up your scanner using the Scanner Wizard, if this is not already done.	Configures OmniPage SE to work with your scanner.
2.	Select Start ➤ Programs ➤ ScanSoft OmniPage SE ➤ OmniPage SE	Opens OmniPage SE on your computer.
3.	Place the document correctly in your scanner.	
4.	Check the three tab buttons to the left of the OmniPage Toolbox. The AutoOCR button should be selected. If not, click on it.	Specifies that you want OmniPage SE to process the document automatically according to the given settings.
5.	From the Get Page drop-down menu, select a scan option for your document: black-and-white, grayscale or color.	Allows you to determine how pictures or colored texts and backgrounds will look in the exported document. Color scanning needs a color scanner.
6.	From the Describe Original drop down menu, check Automatic is selected. For a wide range of documents, this is the best choice.	Configures OmniPage SE to place zones on the page and decide their properties automatically.
7.	From the Export Results drop-down menu, check that Save as File is selected.	This means you will be able to name your export file after you have proofed the document.
8.	Click on Start.	OmniPage SE will start to scan in your document.
9.	The OCR Proofreader appears and invites you to modify words that the program suspects have not been recognized correctly.	The OCR Proofreader operates like a spell checker in a word processing program, but with added OCR-specific features.
10.	Click in the Text Editor. Select Text Editor views one after another, to see how the page appears in each view. Choose the view you want for export.	Each Text Editor view defines a formating level. The view set at saving time is applied to the text in the saved file.
11.	Click Resume to restart proofing. When the message *OCR Proofreading is complete* appears, click on OK.	This ends the OCR Proofreader process. The Save As dialog box will appear.
12.	Choose the location and file type to save your recognized document. Click on OK.	By default, Save and Launch is enabled, so your document will be automatically opened in the word processing program associated with the file type that you selected.
13.	Inspect the document in your word processing program.	You have successfully used OmniPage SE to recognize your document and open it in your target application!

Tip If you suceeded in getting good results from the sample image files, but not from the scanned page, check your scanner installation and settings: in particular brightness and image resolution. See *Input from scanner* for a model of optimum brightness. See also the online Help topics *Setting up your Scanner* and *Scanner Troubleshooting*.

Here is an overview of the processing methods you can use. You will find step-by-step guidance for each of them in the following pages.

Using the OCR Wizard

The OCR Wizard guides you through the selection of settings and commands by asking you questions. It then launches automatic processing. This is a good way to get started if you are new to OmniPage SE.
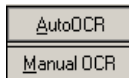
Automatically

The fastest and easiest way to process documents is to let OmniPage SE do it automatically for you. Select settings in the Options dialog box and commands in the AutoOCR toolbar and then click *Start*. It will take each page through the whole process from beginning to end, when possible running in parallel. It will typically auto-zone the pages.

Manually
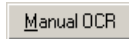
Manual processing gives you more precise control over the way your pages are handled. You can process the document page-by-page with different settings for each page. The program also stops between each step: acquiring images, performing recognition, exporting. This lets you, for instance, draw zones manually or change recognition language(s). You start each step by clicking buttons on the Manual OCR toolbar.

Automatically with manual finishing

You can process a document automatically and view results in the Text Editor. If most pages are in order, but a few have not turned out as expected, you can switch to manual processing to adjust settings and rerecognize just those problem pages.

In other applications

You can use the Direct OCR feature to call on the recognition services of OmniPage SE while working in your usual word-processor or similar application. OmniPage SE automatically links itself to ScanSoft's PaperPort and Pagis document management programs.

At a later time

You can schedule OCR jobs to be performed automatically at a later time, when you may not even be present at your computer. The Add Job Wizard in Schedule OCR allows you to specify settings and a starting time.

PROCESSING DOCUMENTS USING THE OCR WIZARD

The OCR Wizard takes you through six settings panels, guiding you to make settings for your document and then launching automatic processing. Context-sensitive help is available for all Wizard panels. The OCR Wizard can run only when there is no document open in OmniPage SE.

OCR Wizard

Click the OCR Wizard tab in the OmniPage Toolbox and click the Wizard button to see the first wizard screen:
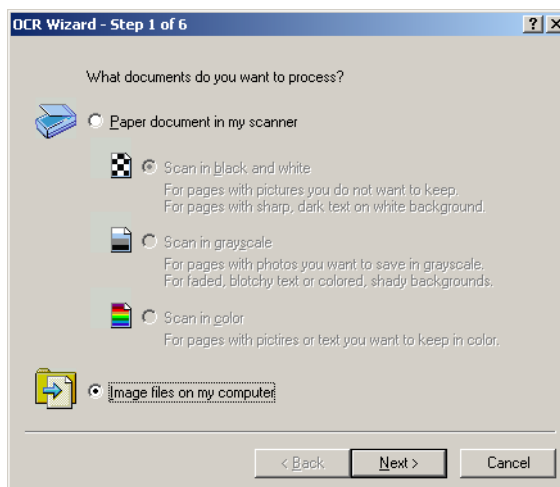
1. The first panel lets you define your document source: scanner or image file. For more information, see the section *Defining the source of page images*. Answer the questions in the first screen and click *Next*.

2. The second panel asks you to describe the layout of the input document, to assist the auto-zoning. For more information, see the section *Describing the layout of the document*.

3. The third panel (shown below) lets you define recognition languages and decide OCR method. Languages with dictionary support have the icon 📖.
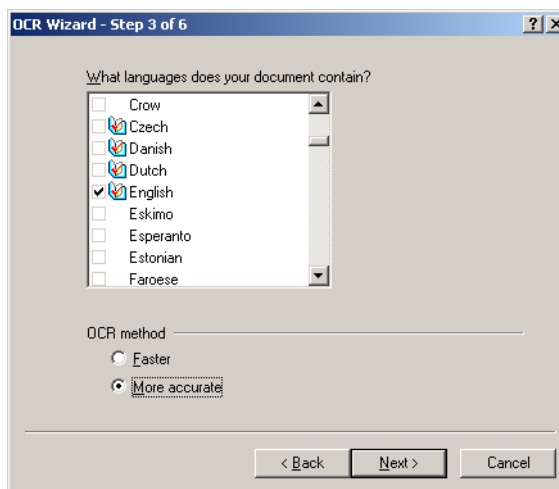
4. The fourth panel lets you define the formatting level to be applied to your document for display and export. See *The editor display and views* in chapter 4 for more information.

5. The fifth panel asks if you want to proofread the text before export. If you choose *Yes* you can also edit the text before saving. You also decide whether to create and use IntelliTrain data during proofing. See chapter 4 for more information. The reference to IntelliTrain does not apply to OmniPage SE.

6. The last panel asks you to define the export choice: saving to file or copying to Clipboard. After setting the choice, click *Finish* to close the Wizard and start the automatic processing.

7. If you requested proofing and the text contains suspect words, the OCR Proofreader™ dialog box will appear. When proofing is finished or closed, recognition results either go directly to the Clipboard, or the Save As dialog box appears so you can specify file export settings.

8. The document remains in OmniPage SE. You can edit recognition results and save it again to other formats. You can change zones manually or change other settings and then use manual processing to rerecognize single pages from the document. You can add pages with automatic or manual processing.

Note The Wizard panels present settings as they were last set in the program. Also, OmniPage SE will remember the settings you make in the OCR Wizard panels and apply them to future automatic or manual processing, until you change them. So, if you have more documents for which your OCR Wizard settings are suitable, just switch to the AutoOCR toolbar and click *Start*.

Note Applicable settings not offered by the OCR Wizard take the values last set in the program. This concerns mainly scanner settings, a user dictionary or a training file. Zone templates cannot be used with the OCR Wizard. If a template file was set when the OCR Wizard starts, it is unloaded and Automatic is set as input description. You cannot export a recognized document as a mail attachment. Please use automatic or manual processing for this.

PROCESSING DOCUMENTS AUTOMATICALLY

Automatic processing provides an efficient way of handling documents, especially larger ones. First you select all settings needed, then you can use the AutoOCR™ toolbar in the OmniPage Toolbox to process a new document from start to finish or to restart and finish processing on an open document.
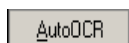
AutoOCR

1. Click the AutoOCR tab in the OmniPage Toolbox to display the AutoOCR toolbar.
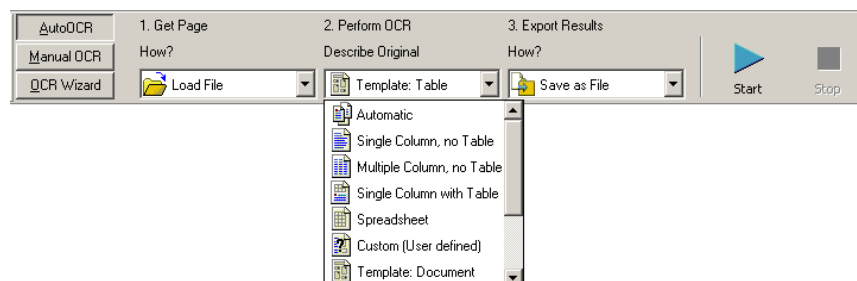
2. Select the desired Get Page command in the drop-down list. You define the document source, which can be from image files or from a scanner. For more detail see the section *Defining the source of page images*.

3. Select a command from the Describe Original drop-down list, as shown above. This guides the program in auto-zoning the pages. You describe the incoming pages or specify a zone template file. For more information on the choices, see the section *Describing the layout of the document*.

4. Select a command from the Export Results drop-down list. You can save the recognized document to file, copy it to Clipboard or send it as a mail attachment. For information on the choices, see chapter 5.

5. Choose Options in the Tools menu and check that settings are appropriate for your document. You can, for instance, specify recognition languages and whether you want to proofread the document or not. See *Settings* at the end of chapter 2.

6. Click *Start* or choose Start in the Process menu. Each page of the document is processed and finished one after the other. The program may perform tasks simultaneously, for instance it may start loading and recognizing a new page as you proofread the previous page.

Command buttons

Start: This lets you begin automatic processing on a new document.

Stop: This lets you interrupt automatic processing. You may do this if you find that some settings need to be changed. Then the *Start* button changes to *Finish.*

The start button takes different values when processing is stopped or finished.

Finish: This appears if processing is incomplete. It lets you:
  ◆ **Finish** processing unfinished pages.
  ◆ **Export** the document, dropping any unrecognized pages.

Additional: This appears if all existing pages are processed and have been exported once. It lets you:
  ◆ **Export** the document again, maybe with changes, to a different file type, name or location, or with a different formatting level.
  ◆ **Add more pages:** from the same source or a different source, with changed or unchanged settings.
  ◆ **Re-process all pages:** Discard all recognition results and rerecognize all pages in the document with different settings. You can specify auto-zoning or a template file.

Tip You may reprocess all pages if an unsuitable setting caused poor results on all pages. An example is incorrect language choice, resulting in almost all words marked suspect during proofing. 'Re-process' lets you perform rerecognition without having to scan or load or rezone all the images again.

PROCESSING DOCUMENTS MANUALLY

Manual processing gives you more precise control over the way your pages are handled. You can process the document page-by-page with different settings for each page. The program also stops between each step: acquiring images, performing recognition, exporting. This lets you, for instance, draw zones manually on each page. You start each step in the process by clicking the buttons on the Manual OCR toolbar.
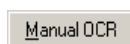
| Manual OCR |

1. Click the Manual OCR tab in the OmniPage Toolbox to display the Manual OCR toolbar.
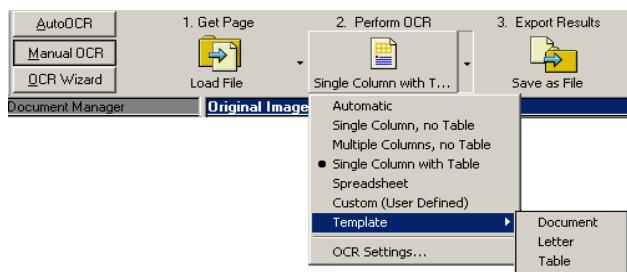
2. Click the icon in the Standard toolbar or Options in the Tools menu to check or make settings in the Options dialog box. See *Settings* at the end of chapter 2.

3. Select the desired value for the Get Page button. You define the document source, which can be from image files or from a scanner. Access the scanner settings dialog box and make settings as desired. For more detail see the section *Defining the source of page images*.

4. Click the Get Page button. This either brings up the Load File dialog box allowing you to name images files, or initiates scanning. The result is one or more images displayed in the Document Manager and one in the Original Image area.

5. Now you can manually draw and modify zones on one or more images and assign properties. Status bar buttons let you move to other pages. Any image without zones will be auto-zoned when recognition is requested. For guidance, see the section *Manual zoning*.

6. Select a value for the Perform OCR button. You describe the layout of the incoming pages. This value has an influence if auto-zoning runs on any pages. You can also select a template to have its zones placed on the current page. For more detail see the sections *Describing the layout of the document* and *Using zone templates.*

7. Click the Perform OCR button to have the current page recognized. To have selected pages recognized, make a multiple selection in the Document Manager (see *Managing documents* in chapter 2) and then click the Perform OCR button.

8. The Zoning Instructions dialog box appears, unless you disabled it.
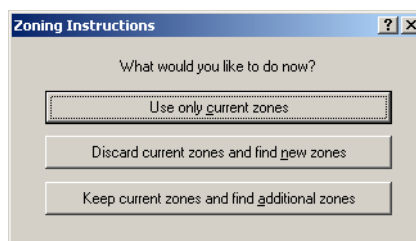
When you choose one of its options, recognition starts.

9. If you requested proofing, the OCR Proofreader dialog box displays suspect words one after the other from the recognized page(s). You can proof and edit the recognized text. See *Proofreading OCR results* in chapter 4.

10. Continue loading pages, performing OCR, editing and proofing as desired.

11. Select a value for the Export Results button. You can save the recognized document to file (including as an OmniPage Document), copy it to Clipboard or send it as a mail attachment. You can save the document more than once; see *Saving recognition results* in chapter 5.

Note If you deselect 'Find zones in addition to template/current zones' in the Process panel of the Options dialog box, the Zone Instructions dialog box will not appear and recognition will always run with current zones only.

Processing a document automatically and finishing it manually

When you have a large document with only a few pages needing special attention, you do not have to manually process the whole document. You can process it automatically and view results in the Text Editor. You can determine which pages are in order, and which need different settings or some manual zoning. Then you can switch to manual processing to adjust settings and zones and rerecognize just those pages.

1. Prepare the document and perform automatic processing, as already described.

2. If you close or finish proofing you will be invited to save the document. This is recommended, even if it is not in its final form.

3. Select a page needing rezoning or changed settings and click the Manual OCR tab at the left of the OmniPage Toolbox.

4. Delete or modify the existing zones in the Original Image area. You can also load a template to let its zones replace existing ones. Draw new zones as desired. See *Manual zoning*.

5. Change other settings as required for the current page. See *Settings* at the end of chapter 2.

6. Click the Perform OCR button to rerecognize the current page. Confirm that the previous recognition results should be overwritten. The Zoning Instructions dialog box will appear, unless disabled.

7. To rerecognize more than one page, select the required pages in the Document Manager before clicking the Perform OCR button.

8. When all pages have been rerecognized with acceptable results, save the document again.

PROCESSING FROM OTHER APPLICATIONS

You can use the Direct OCR feature to call on the recognition services of OmniPage SE while you work in your usual word-processor or other application. First you must establish the direct connection with the application. Then, two items in its File Menu open the door to OCR facilities.

How to set up Direct OCR

1. Start the application you want connected to OmniPage SE. Start OmniPage SE, open the Options dialog box at the Direct OCR panel and select 'Enable Direct OCR'.

2. The Unregistered panel displays running or previously registered applications. Select the desired one(s) and click Add. You can browse for an unlisted application. Select the process options as desired, to function as preferences.

How to use Direct OCR

1. Open your registered application and work in a document. To acquire recognition results from scanned pages, place them correctly in the scanner.

2. Use the File Menu item Acquire Text Settings... to specify settings to be used during recognition. Any settings not offered take their values from those last used in OmniPage SE. Settings changed for Direct OCR are also changed in OmniPage SE.

3. Use the File Menu item Acquire Text to acquire images from scanner or file.

4. If you selected 'Draw zones automatically' in the Direct OCR panel of the Options dialog box, or under Acquire Text Settings..., recognition proceeds immediately.

5. If 'Draw zones automatically' is not selected, each page image will be presented to you, allowing you to draw zones manually. Click the Perform OCR button to start recognition.

6. If proofing was specified, this follows recognition. Then the recognized text is placed at the cursor position in your application, with the formatting level specified by Acquire Text Settings... .

Note If OmniPage SE is running when Direct OCR is called from a target application, a second instance of OmniPage SE is launched.

How to use OmniPage SE with your PaperPort software

PaperPort® is a paper management software product from ScanSoft. It lets you link pages with suitable applications. Pages can contain pictures, text or both. If PaperPort exists on a computer when OmniPage SE is installed, its OCR services become available and amplify the power of PaperPort. You can choose an OCR program by right clicking on a text applications PaperPort link, selecting Preferences and then selecting OmniPage SE as the OCR package. OCR settings can be specified, as with Direct OCR.
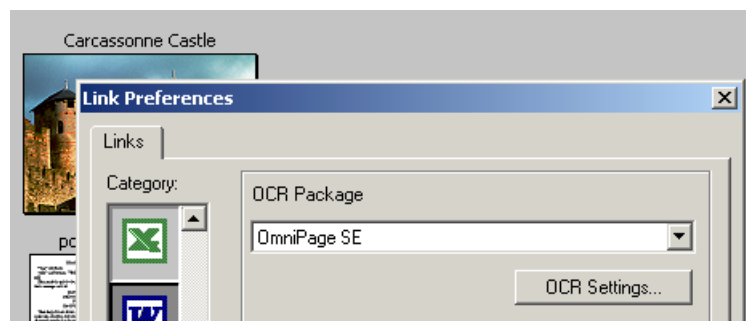
Here OmniPage SE has been selected as the OCR package for MS Word 2000. Then you can drag page images from the PaperPort desktop onto the MS Word link on the PaperPort. While the text is being recognized, only a progress monitor is displayed. OmniPage SE's manual zoning window or proofing facility will appear if requested. The recognition results are placed in a new unnamed document in the target application.

PROCESSING DOCUMENTS WITH SCHEDULE OCR

You can schedule OCR jobs to be performed automatically at any time within the following 24 hours. Each job handles one document. The document pages can come from a scanner with an ADF or from image files. You do not have to be present at your computer at job start time, nor does OmniPage SE have to be running. It does not matter if your computer is turned off after the job is set up, so long as it is running at job start time. If you are scanning pages, your scanner must be functioning at job start time, with the pages loaded in the ADF. Here is how to set up a job:

1. Click Schedule OCR in the Process menu or in the Windows Start menu: select Programs▶ScanSoft▶OmniPage SE▶Schedule OCR.

2. The Schedule OCR dialog box appears. Click Add Job... to get the Add Job Wizard. It takes you through six panels, similar to the OCR Wizard.

3. In the first panel you define image source. An additional feature lets you process all supported image files in a defined folder.

4. The next three panels are similar to those in the OCR Wizard, but you can also specify a user dictionary. In OmniPage Pro 11 you can specify a training file and/or run IntelliTrain. These are not available in OmniPage SE.

5. The fifth panel lets you specify an export file name, type, location and a file separation choice.

6. The last panel lets you define the job start time, retain or delete input files after processing and specify use of a log file to note job completion and any problems encountered. Click Finish to close the Wizard.

Note The Schedule OCR dialog box lists all jobs, with status Waiting, Running, Error or Complete. Use Modify Job... to change settings for a waiting job. You can modify and reuse finished jobs to process new jobs needing similar settings. You can delete completed jobs when they are no longer needed.

For more information, please see *Scheduling OCR* in the online Help.

DEFINING THE SOURCE OF PAGE IMAGES

There are two possible image sources: from image files and from a scanner. There are two main types of scanners: flatbed or sheetfed. A scanner may have a built-in or added Automatic Document Feeder (ADF), which makes it easier to scan multi-page documents. The images from scanned documents can be input directly into OmniPage SE or may be saved with the scanner's own software to an image file, which OmniPage SE can later open.

Input from image files

You can create image files from your own scanner, or receive them by e-mail or as fax files. OmniPage SE can open a wide range of image file types; see *Supported file types* in chapter 6. Image files are specified in the Load File dialog box. This appears when you start automatic processing. In manual processing, click the Load File button or use the Process menu. The lower part of the dialog box provides advanced settings, and can be shown or hidden. Here it is displayed.

This is the current folder.

Use Shift+ clicks or Ctrl+clicks to place more than one file in the File name text box.

Specify the file type(s) you want listed.

This can be used for multipage TIFF and DCX files.

This is a blank image file for the saving option: "New file for each blank page".

Load File		? x
Look in: ScanSoft		

OmniPagePro11.0 · 1 letter.tif · pentax.tif
0001.tif · 5 WE langs.tif · selling.tif
0002.tif · car.tif · tables nowrap.tif
0003.tif · differ 22.tif · telecom.tif
0004.tif · gauge.tif · time-sheet.tif
0005.tif · ordinal1-45.tif · warn.tif

File name: ordinal1-45.tif — Open

Files of type: Image files [*.tif; *.pcx; *.dcx; *.bmp; *.jpg; *. — Cancel / << Basic

Page range: e.g.: 1, 2, 5-10

Selected files:

D:\Program Files\ScanSoft\time-sheet.tif
D:\Program Files\ScanSoft\warn.tif
D:\Program Files\ScanSoft\gauge.tif
C:\bitmap\Book2.tif
C:\bitmap\DropCap2.tif
C:\Image files\blank.tif

Add / Add All / Remove / Remove All

Click Advanced to open the lower panel and Basic to close it.

Use this to add files one by one from different folders and to control file order precisely.

Normally the Add button places each file at the bottom of the file list. To place a file at a different location, highlight a file in the list. The new file will be added immediately below the lowest highlighted file.

Input from scanner

You must have a functioning, supported scanner correctly installed with OmniPage SE. See *Setting up your scanner with OmniPage SE* in chapter 1 for more information. You have a choice of scanning modes. In making your choice, there are two main considerations:

◆ Which type of output do you want in your export document?
◆ Which mode will yield best OCR accuracy?

Scan black and white
Select this to scan in black-and-white. This is not suitable if you want color in your output document, nor if you want pictures to look like so-called 'black-and-white' photographs: they need grayscale scanning. For best OCR accuracy, use this for crisp black texts on a white or light background. Black-and-white images can be scanned and handled quicker than others and occupy less disk space.
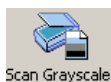
Scan grayscale
Select this to use grayscale scanning. Choose this to keep 'black-and-white' photographs in the output document. For best OCR accuracy, use this for pages with varying or low contrast (not much difference between light and dark) and with text on colored or shaded backgrounds.

Scan color
Select this to scan in color. Available only with color scanners. Choose this if you want colored graphics, texts or backgrounds in the output document. For OCR accuracy, it offers no more benefit than grayscale scanning (for a given resolution), but will require much more time, memory resources and disk space.

Brightness and contrast

Good brightness and contrast settings play an important role in OCR accuracy. Set these in the Scanner panel of the Options dialog box. The diagram illustrates an optimum brightness setting. After loading an image, check its appearance. If characters are thick and touching, lighten the brightness. If characters are thin and broken, darken it. Then rescan the page.

extensive research — Unsuitable

extensive research

extensive research — Tolerable

extensive research

extensive research — Good

extensive research

extensive research — Best

extensive research — Good

extensive research

extensive research — Tolerable

extensive research

extensive research — Unsuitable

Scanning with an ADF

The best way to scan multi-page documents is with an Automatic Document Feeder (ADF). Simply load pages in the correct order into the ADF. Place blank pages if you want to save your document to multiple output files using the 'Create a new file at each blank page' option. See *Saving to file* in chapter 5.

If you have a document longer than the capacity of your ADF, select 'Automatically prompt for more pages' in the Process panel of the Options dialog box. Then a dialog box lets you add further page batches and signal when all pages are scanned.

You can scan double-sided documents with an ADF. A duplex scanner will manage this automatically. For non-duplex scanners, select 'Scan double-sided pages' in the Scanner panel of the Options dialog box. Then you can scan the document in just a few passes, with even pages grouped together and odd pages also grouped. OmniPage SE will merge the pages for you.

Scanning long documents without an ADF

You can scan multi-page documents efficiently from a flatbed scanner, even without an ADF. Select 'Automatically scan pages' in the Scanner panel of the Options dialog box, and define a pause value in seconds. Then the scanner will make scanning passes automatically, pausing between each scan by the defined number of seconds, giving you time to place the next page. A dialog box allows you finish the pause early or request a longer pause and to specify when the last page is scanned.

DESCRIBING THE LAYOUT OF THE DOCUMENT

Before starting recognition you are requested to describe the layout of the incoming pages to assist the auto-zoning process. When you use the OCR Wizard, auto-zoning always runs. When you do automatic processing, auto-zoning always runs unless you specify a template to be used on its own. When you do manual processing, auto-zoning sometimes runs. See online Help for more detail.

Here are your input description choices:

Automatic
Choose this to let the program make all auto-zoning decisions. It decides whether text is in columns or not, whether an item is a graphic or text to be recognized and whether to place tables or not. Choose Automatic if your document contains pages with different or unknown layouts. Choose it for a page with multiple columns and a table, and for any pages with more than one table.

Single column, no table

Choose this setting if your pages contain only one column of text and no table. Business letters or pages from a book are normally like this. Choose it also for a page with words or numbers arranged in columns if you do not want these placed in a table or decolumnized or treated as separate columns. Graphics may be detected.

Multiple columns, no table

Choose this if some of your pages contain text in columns and you want this decolumnized or kept in separate columns, similar to the original layout. Columns can be retained in the output document, either with frames (if True Page is set) or without frames (if Retain Flowing Columns is set). If tabular data is encountered, it is likely to be treated as flowing text. Graphics may be detected.

Single column with table

Choose this if your page contains only one column of text and a table. Auto-zoning will not look for columns but will try to find a table and place it in a grid in the Text Editor. You can later specify whether to export it in a grid or as tab separated text columns. Graphics may be detected.

Spreadsheet

Choose this if your whole page consists of a table which you want to export to a spreadsheet program, or have treated as single table. No flowing text or graphics zones will be detected.

Custom

Choose this for maximum control over auto-zoning. You can prevent or encourage the detection of columns, graphics and tables. Make your settings in the Custom Layout panel of the Options dialog box.

Template

Choose a zone template file if you wish to have its zones and properties applied to all acquired pages from now on. In manual processing the template zones are also applied to the current page, replacing any existing zones. Other zones are permitted in addition to template zones. For more detail, see the section *Using zone templates*.

If auto-zoning yielded unexpected recognition results, use manual processing to rezone individual pages and rerecognize them.

MANUAL ZONING

Zones define areas on the page to be processed. Zones are rectangular or irregular (with sides formed by vertical and horizontal lines). Zones cannot overlap. They have a zone number in the top left corner and a zone type icon top right. Click in a zone to select it. Use Shift+clicks for a multiple selection. Current and selected zones are shaded. Click outside a zone to remove the selection. Zones appear on an original image in the following cases:

◆ The page has been recognized.
◆ A zone template file was specified in manual processing while the page was current.
◆ You have drawn manual zones on the image.

Working with zones

The Image toolbar provides zone editing tools. One is always selected. When you no longer want the service of a tool, click a different tool. Normally this will be the Draw Rectangular Zones tool.

Draw rectangular zones
Click this and drag the cursor to define rectangular zones. The new zone takes its properties from the last drawn or selected zone. You can also move or resize existing zones when this tool is active.

Draw irregular zones
Click this for a tool allowing you to draw irregular zones. Click and drag to draw a single line. Repeat until only one line remains undrawn. Double-click to close the shape. Irregular zones snap to rectangles if you set them as table type zones. You can also move or resize existing zones when this tool is active.

Add to zone
Click this to make irregular additions to an existing zone or combine separate zones into one. You cannot move or resize existing zones when this tool is active. You cannot use this with a table type zone.

Subtract from zone

Click this to subtract irregular parts from an existing zone or split a zone into smaller ones. You cannot move or resize existing zones when this tool is active. You cannot use this with a table type zone.

Reorder zones

Click this for the zone reordering tool. Then click in zones in the desired reading order. For your order to be respected, choose 'Use current zones only' and avoid having multiple-column or auto-detect zones types on the page.

Zone properties

Click this for the Zone Properties dialog box. This lets you define zone type and content for the currently selected zone(s) on the page. You can also do this from a zone's shortcut menu. See the next section.

Zone properties

Each zone has a zone type. Zones containing text can also have a zone contents setting: alphanumeric or numeric. The zone type and zone contents together constitute the zone properties. Right-click in a zone for a shortcut menu allowing you to change the zone's properties. Select multiple zones to change their properties in one move. The zone properties button in the Image toolbar can be used for the same purpose. The following types are available:

Single-column flowing text zone

Use this to have zone contents treated as flowing text, without columns being found.

Multiple-column flowing text zone

Use this to have zone contents treated as flowing text. The program will try to detect columns inside the zone. Text will be decolumnized or retained in columns, depending on the Text Editor view. During recognition, a multi-column zone may be replaced by separate zones for each column. To do this, auto-zoning must run, which may also result in changed zone order.

Table zone

Use this to have the zone contents treated as a table. Table grids can be automatically detected, or placed manually as described in the next section. Table zones must be rectangular. The Text Editor displays the table in an editable grid. You can choose whether to export tables in grids or in columns separated by tabs.

Auto-detect zone

Use this to let the program decide the zone type. To do this, auto-zoning runs, which may also result in changed zone order on the page. After recognition you can see the type that was applied. If you use an auto-detect zone to cover a page area with varied contents, the program may replace the auto-detect zone with a number of smaller zones.

Graphic zone

Use this to enclose a picture, diagram, drawing, signature or anything you want transferred to the Text Editor as an embedded image, and not as recognized text. A graphics zone has a green border. Embedded images can be exported with the document to target applications supporting graphics.

Ignore zone

Use this to define a page area you do not want in the Text Editor. Auto-zoning will not place zones here. To exclude a given page area from many pages (for example a header or page numbers), place ignore zones in a template and select 'Find zones in addition to template/current zones' in the Process panel of the Options dialog box.

Zone contents

This is available for zone types containing text. Alphanumeric contents validates all characters needed for your language choice. Recognition results from a numeric zone will contain only numbers and number-related punctuation. No letters will be placed.

Note   Right-click outside a zone for a shortcut menu tailored for the whole image. It allows you to zoom in or out or rotate the image. When an image is rotated, all zones on it are deleted.

TABLE GRIDS IN THE IMAGE
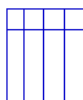
After automatic processing you may see table zones placed on a page. They are denoted with a table zone icon in the top right corner of the zone. To change a zone to or from a table zone, use its shortcut menu. You can also draw a table type zone. If there is already a table zone on the page, select it, then draw the new rectangular zone. It will inherit the table type. Otherwise draw a rectangular zone and use its shortcut menu to change it to a table type.

You draw or move table dividers to determine where gridlines will appear when the table is placed in the Text Editor. You can use the Add or Subtract tools to enlarge or reduce a table zone, but it must remain rectangular. You can do this to discard unneeded columns or rows from a table.

The five table handling tools on the Imaging toolbar become active if the current page contains a table type zone. Use them as follows:

Move row or column dividers

Click the tool and move the cursor to the divider to be moved. It displays a double-headed arrow. Drag the border as desired. You cannot drag it beyond its neighbor. Avoid placing dividers so they overlap one another or cut through text. Press the Ctrl key as you drag a column divider, to move it in the current row only.

Insert column dividers

Click the tool then click at the location in a table zone where you want to place a column divider. Press the Ctrl key as you click to place the divider in the current row only.

Insert row dividers

Click the tool then click at the location in a table zone where you want to place a row divider. Avoid placing a divider on top of another one or so it cuts through text.

Remove column or row dividers

Click the tool then click on a single divider you want to delete. Do this if a divider is wrongly located, or if you want to change the appearance of the table in the final document. For example, you can place two columns of data in a single column by deleting the divider between the columns.

Remove/replace all dividers

Click this tool and click inside a table zone. Its dividers will all disappear. Click again to have dividers automatically (re)detected. Divider placement usually occurs during recognition; clicking twice with this tool lets you see and edit the dividers before recognition.

USING ZONE TEMPLATES

A template is a set of zones, their properties and reading order, stored in a file. A zone template file can be loaded to have template zones used during recognition. Load a template file in the Perform OCR drop-down list or from the Tools menu.

When you load a template with the **Manual OCR toolbar**, its zones appear immediately on the current page, replacing any already there. Existing pages are not affected. The template zones are placed on all further acquired pages until the template is unloaded. You can modify the template zones and add new zones before performing recognition.

When you load a template with the **AutoOCR toolbar**, it does not affect the current or existing pages. The template zones are placed on all further acquired pages until the template is unloaded. The Process panel of the Options dialog box presents the option 'Find zones in addition to template/current zones'. If this is turned on during automatic processing, auto-zoning will run on page areas outside the template zones.
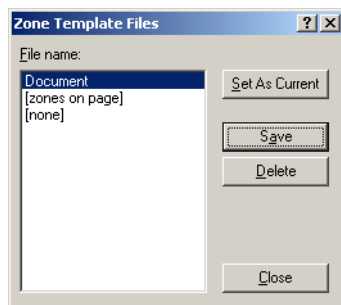
How to save a zone template

Prepare zones on a page. Check their locations, properties and reading order. Click Zone Template File... in the Tools menu. In the dialog box, select [zones on page] and click Save.

How to modify a zone template

Load the template and acquire a suitable image with manual processing. The template zones appear. Modify the zones and/or properties as desired. Open the Zone Template File dialog box. The current template is selected. Click Save and then Close.

How to unload a template

Select a non-template setting for layout description in the Perform OCR drop-down list. The template zones are not removed from the current or existing pages, but template zones will no longer be used for future processing. You can also open the Zone Template Files dialog box, select [none] and click the Set As Current button. In this case, the layout description setting returns to Automatic.

How to replace one template with another

Select a different template in the Perform OCR drop-down list, or open the Zone Template Files dialog box, select the desired template and click the Set As Current button. When the AutoOCR toolbar is active, no existing zones are changed and the new template is used for future processing. When the Manual OCR toolbar is active, zones from the new template are applied to the current page, replacing any existing zones.

How to delete a template file

Open the Zone Template Files dialog box. Select a template and click the Delete button. Zones already placed by this template are not removed.

Tip Templates accept ignore and auto-detect type zones. A template can therefore be useful to define which parts of the page to read, and which parts to ignore.

Note Auto-detect type zones from a template may be replaced during recognition by smaller ones; specific zone types will be assigned to these zones. Multi-column zones may also be split into smaller single-column zones, one for each detected column.

Note Templates and the additional auto-zoning feature are available in Schedule OCR and Direct OCR, but not in the OCR Wizard.

4 Proofing and editing

Recognition results are placed in the Text Editor. This newly developed WYSIWYG (What You See Is What You Get) editor offers the following features, detailed in this chapter:

◆ Proofreading OCR results

◆ Checking recognized text against original (Verifying text)

◆ User dictionaries

◆ IntelliTrain

◆ The editor display and views

◆ Text and image editing

◆ Reading text aloud

◆ Page outline

The Text Editor offers four views for displaying its pages. You can switch freely from one view to another. These provide different levels of formatting. The views are:

No Formatting view
This displays plain decolumnized text in a single font and font size.

Retain Fonts and Paragraphs view
This displays decolumnized text with font and paragraph styling.

True Page view
This view tries to conserve as much of the formatting of the original document as possible. Character and paragraph styling is retained. All page elements, including columns, are placed in frames.

Retain Flowing Columns view
This view is identical to True Page view, except that the reading order of zones is shown by arrows. This view's difference from True Page relates mainly to export, as explained in the section *Preparing recognition results for export* in chapter 5.

PROOFREADING OCR RESULTS

After a page is recognized, the recognition results appear in the Text Editor. Proofreading starts automatically if that was requested in the Proofing panel of the Options dialog box or in the OCR Wizard. You can start proofing manually any time the program is not busy. Work as follows:

1. Click the Proofread OCR button in the Standard toolbar, or choose Proofread OCR... in the Tools menu.

2. Proofing starts from the beginning of the document, but skips text already proofed. If a suspected error is detected, the OCR Proofreader dialog box displays the error and a picture of how it originally looked in the image.

This tells why the word is suspected.

This is what OmniPage SE thought the word was.

The image of the suspect word is highlighted.

This window shows the relevant part of the original image. Click inside it to enlarge or reduce the display.

Drag a corner or the bottom of the dialog box to resize it.

OCR Proofreader

Not found in dictionaries: **fieelance**

Change to: freelance

Suggestion: freelance / filename

User dictionary: Custom

Ignore / Ignore All / Change / Change All / Add / Close

Assistant editor Michael

a *freelance* reporter/resear

3. If the recognized word is correct, click Ignore or Ignore All to move to the next suspect word. Click Add to add it to the current user dictionary and move to the next suspect word.

4. If the recognized word is not correct, edit the word in the Change to edit box, or type in the desired word or select a dictionary suggestion. Click Change or Change All to implement the change and move to the next suspect word. Click Add to add the word in the Change to edit box to the current user dictionary and move to the next suspect word.

5. Color markers are removed from words in the Text Editor as they are proofread. You can switch to the Text Editor during proofing to make corrections there. Use the Resume button to restart proofing. Click Close to stop proofreading before the end of the document is reached.

Note A page is marked with the proofed icons ✔ and 👓 in the Document Manager if proofing ran to the end of the page.

CHECKING RECOGNIZED TEXT AGAINST ORIGINAL

After performing OCR, you can compare any part of the recognized text against the corresponding part of the original image, to verify that the text was recognized correctly. Work as follows:

1. Double-click any word in the Text Editor or select a word and choose Verify Text in the Tools menu. The Verify Text window opens and shows a picture of the original word and its surrounding area. Modify the word in the Text Editor as necessary.

2. Click inside the window to enlarge or reduce the picture. The picture is enlarged on the first two clicks and reduced on the next two clicks.

This is the original image of the word you are verifying.

Close button

asks, "How import at the **story** on the] is a story that ha

This is the word you double-clicked in the Text Editor.

the teacher that the story on the front of [the magazine] is a story that happened
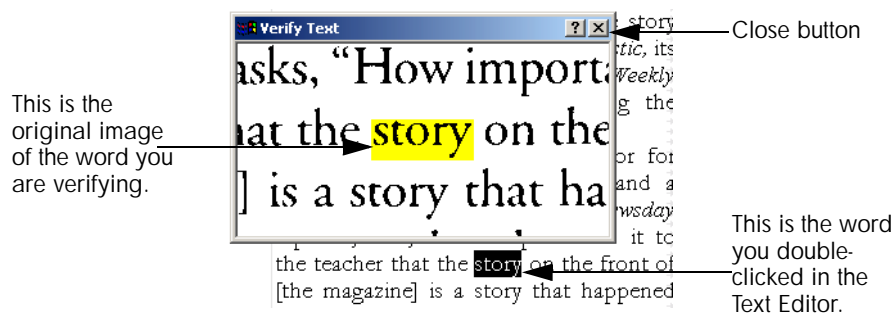
3. Continue double-clicking words that you want to verify, and correcting them as necessary. The display changes as you select new words.

4. Click the Close button to close the verifier window.

Tip You should proofread and verify texts before doing large-scale editing. If you cut and paste large blocks of text, the links between text and image may be disturbed.

Tip OmniPage Pro 11's Text-to-Speech facility can read recognized text aloud as another way of verifying text. You can hear the text letter-by-letter, word-by-word, line-by-line, sentence-by-sentence or in whole pages. See *Reading text aloud.* This is not available in OmniPage SE.

USER DICTIONARIES

The program has built-in dictionaries for many languages. These assist during recognition and may offer suggestions during proofing. They can be supplemented by user dictionaries. You can save any number of user dictionaries, but only one can be loaded at a time. Your user dictionaries from Microsoft Word are also available; a dictionary called Custom is the default user dictionary for Microsoft Word.

Starting a user dictionary
Click Add in the OCR Proofreader dialog box with no user dictionary loaded or open the User Dictionary Files dialog box from the Tools menu and click New. You will be asked to name the dictionary immediately.

Loading or unloading a user dictionary
Do this from the OCR panel of the Options dialog box or from the User Dictionary Files dialog box. Select a dictionary file to load it or [none] to unload a user dictionary.

Editing a user dictionary
Add words by loading a user dictionary and then clicking Add in the OCR Proofreader dialog box. You can add and delete words in the User Dictionary Files dialog box.

Tip While editing a user dictionary, you can import a word list from a text file to add words to the dictionary quickly.

INTELLITRAIN

IntelliTrain is a newly developed and automated form of training. It takes input from the corrections you make during proofing. When you make a change, it remembers the character shape involved, and your proofing change. It searches other similar character shapes in the document, especially in suspect words. It assesses whether to apply the user correction or not.

IntelliTrain and training files are not supported in OmniPage SE. This section applies only to OmniPage Pro 11. Any training data in an OPD file will be ignored when it is opened in OmniPage SE.

You can turn IntelliTrain on or off in the OCR panel of the Options dialog box. It is useful for uniformly degraded documents or when an unusual typeface is used throughout a document. IntelliTrain will be less useful for texts with random distortions. Here is an example, based on the letter "g", which can be printed in different ways:

Rugged···Rugged···**Rugged**··Rugged

The first two examples do not need IntelliTrain, because both shapes are normal for the letter "g" and the program can handle them. The third example could benefit from IntelliTrain because the shape of "g" is unusual, and all instances of "g" in the text are likely to look like this. The fourth example is not good for IntelliTrain, because the first "g" is poorly printed, and this shape is unlikely to appear again in the document.

The following shows how IntelliTrain works, using the original image. Our example involves the letters **c** and **e**. With some typefaces and scanning settings, the horizontal line in **e** can become very thin, leading to OCR errors that IntelliTrain can repair.

OmniPage Pro read this as **bcnefit**.

You changed it during proofing to **benefit**.

Original Image

well-heeled competition, saying the in
duction of *Time for Kids* into the ma
will ultimately benefit his comp:
"Right now, they've decided they war
go out and brand little readers into *T*
readers," Goddy asserts. "If you are in
magazine business for schools and all
sudden there are a million more subsc
tions floating around out there, and y
long-term business—like ours—is c
room magazines, that's going to be g
for us. Whenever they get tired of spe
ing all that money to support that ef:

IntelliTrain remembers this shape and the rule:

e This is not c.
This is e.

IntelliTrain changes:

thcrc to **there**
likc to **like**
Whcncvcr to **Whenever**
etc.

IntelliTrain remembers the training data it collects, and you can save this to a training file for future use with similar documents. If you want to be prompted to save your unsaved training data when you close the document, select that option in the Proofing panel of the Options dialog box. Unsaved training data is stored in an OmniPage Document.

Saving training to file, loading, editing and unloading training files are all done in the Training Files dialog box. Open this from the Proofing panel of the Options dialog box or the Tools menu.

Select this, click Save and type in a name to save a new training file.

Select this to unload a training file.

Training Files

File name:

German
Pg watch
[unsaved]
[none]

Set As Current
Edit
Save
Delete
Close
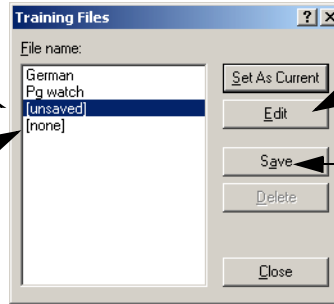
Click this to edit the selected training file (see below).

Use this also to save new training into a loaded training file. It is listed as:
File name [modified]

Unsaved training can be edited in the Edit Training dialog box, an asterisk is displayed in the title bar in place of a training file name. It remains unsaved when you close the dialog box.

A training file can be also edited; its name appears in the title bar. If it has unsaved training added to it, an asterisk appears after its name. Both the unsaved and the modified training are saved when you close the dialog box.

The dialog box displays frames containing a character shape and an OCR solution assigned to that shape. Click a frame to select it. Then you can delete it with the Delete key, or change the assignation. Use arrow keys to move to the next or previous frame.

You are editing your unsaved training.

This frame is grayed. It has been deleted. To undelete it, select it again and press the Delete key. Characters marked as deleted are really deleted when you close the dialog box.

Training Files

File name:

Französisch
Deutsch
Pg watch
[unsaved]
[none]

Set As Current
Edit

Edit Training *
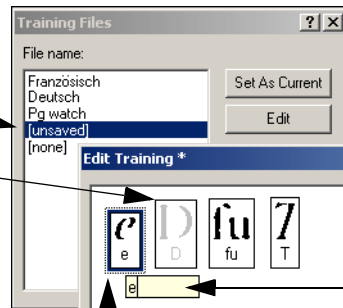
Double-click a frame or press Enter to change its OCR solution. Enter the new solution in the text box that appears and press Enter. Changed assignations appear in red.

This frame is selected. The top part shows the shape from the image. The bottom part shows the assigned OCR solution.

THE EDITOR DISPLAY AND VIEWS

The editor displays recognized texts and can mark words that were suspected during recognition. Marking is done with a wavy underline; red underlines for words not found in a dictionary (this applies only to languages with dictionary support) and blue underlines for words containing suspect or reject characters. These markers can be shown or hidden as selected in the Text Editor panel of the Options dialog box. You can also show or hide non-printing characters and header/footer indicators. The Text Editor panel also lets you define a unit of measurement for the program and a word wrap setting for use in all Text Editor views except No Formatting view.

Here are the main differences between the views:

No Formatting view
This displays plain decolumnized left-aligned text in a single font and font size, with the same line breaks as in the original document. Most formatting buttons and dialog boxes are disabled. Rulers are not displayed. You may find this view convenient for verifying and editing the text.

Retain Fonts and Paragraphs view
This displays decolumnized text with font and paragraph styling. The horizontal ruler is displayed. You may find this view convenient for verifying, editing and modifying the text together with its styling.

True Page view
This view tries to conserve as much of the formatting of the original document as possible. Character and paragraph styling is retained. All page elements, including columns, are placed in frames. It may be more difficult to verify and edit text in this view; you may need to scroll within a frame to see all the frame contents. A row of arrows denote contents extending beyond frame borders.

Retain Flowing Columns view
This view is identical to True Page view, except that the reading order of zones is shown by arrows. This view differs from True Page during export, see the section *Preparing recognition results for export* in chapter 5.

Select a view with the four buttons at the bottom left of the Text Editor or from the View menu. Graphics and tables can appear in all four views.

Text and image editing

This is a WYSIWYG Text Editor, providing many editing facilities. These work very similarly to those in leading word processors.

Editing character attributes
In all views except No Formatting view, you can change the font type, size and attributes (bold, italic, underlined) for selected text. Use the Formatting toolbar or the Font dialog box from the Format menu. The latter also offers subscripts, superscripts and colored text or backgrounds.

In No Formatting view you can use the Formatting toolbar to specify one font type and size to be applied to the whole document. This is not transferred to other views; their previous settings are restored.

Open the Font Matching dialog box from the OCR panel of the Options dialog box to specify which fonts to use for texts entering the Text Editor.

Editing paragraph attributes
In all views except No Formatting view, you can change the alignment of selected paragraphs and apply bulleting to paragraphs. Use the Formatting toolbar or the Paragraph dialog box from the Format menu. The latter allows you to modify indents, line spacing and spacing between paragraphs. The Text Editor's horizontal ruler lets you define indent and tab positions easily. Advanced tab settings are done in the Tabs dialog box from the Format menu.

Paragraph styles
Paragraph styles are auto-detected during recognition. A list of styles is built up and presented in a selection box on the left of the Formatting toolbar. Use this to assign a style to selected paragraphs. Use the Style dialog box from the Format menu to rename or modify a style and to define a new style. When you save a document to file, you can choose whether to export the paragraph styles with the document or not. This is valid only if the target application supports paragraph styles.

Graphics

You can edit the contents of a selected graphic zone if you have an image editor in your computer. Click Edit Picture in the Tools menu. This will activate the image editor associated with BMP files in your Windows system, and load the graphic. Edit the graphic, then close the editor to have it reembedded in OmniPage SE's Text Editor. Do not change the graphic's size, resolution or type, because this will prevent the reembedding.

Tables

Tables are displayed in the Text Editor in grids. Move the cursor into a table area. It changes appearance, allowing you to move gridlines. You can also use the Text Editor's rulers to modify a table. Modify the placement of text in table cells with the alignment buttons in the Formatting toolbar and the tab controls in the ruler. When saving the document to file, you can choose whether to have the tables exported in grids or as tab separated columns.

READING TEXT ALOUD

The Text-to-Speech facility is enabled or disabled with the Tools menu item Speech Mode or with the F5 key. A second menu item Speech Settings... allows you to select a voice (for example, male or female for a given language), a reading speed and the volume.

This speech facility is designed for the visually impaired, but it can also be useful to anyone during text checking and verification. The speaking is controlled by movements of the insertion point in the Text Editor which can be mouse or keyboard driven.

The Text-to-Speech facility is not included in OmniPage SE. It is available in OmniPage Pro 11.

To hear text:	Use these keys:
One character at a time, forward or back	Right or left arrow. Letter, number or punctuation names are spoken.
Current word	Ctrl + Numpad 1
One word to the right	Ctrl + right arrow *
One word to the left	Ctrl + left arrow *
A single line	Place the insertion point in the line
Next line	Down arrow
Previous line	Up arrow
Current sentence	Ctrl + Numpad 2
From insertion point to end of sentence	Ctrl + Numpad 6
From start of sentence to insertion point	Ctrl + Numpad 4
Current page	Ctrl + Numpad 3
From top of current page to insertion point	Ctrl + Home
From insertion point to end of current page	Ctrl + End
Previous, next or any page	Ctrl + PgUp, PgDown or navigation buttons
Typed characters	Each typed character is pronounced, one by one, including punctuation.

* If the cursor is in the middle of a word, you will first hear a word fragment, but from the second keystroke you will hear whole words.

The three basic speech keys are grouped together on the numeric keypad.
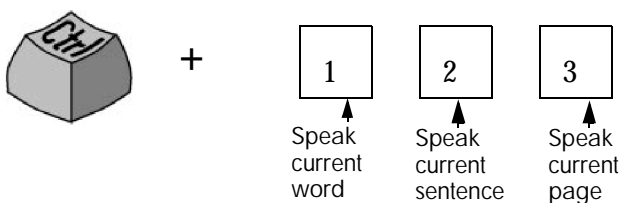
You also have the following keyboard controls:

To do this:	Use this:
Pause/Resume	Ctrl + Numpad 5
Set speed higher	Ctrl + Numpad +
Set speed lower	Ctrl + Numpad -
Restore speed	Ctrl + Numpad *

It is planned to provide speech programs for the following languages: English, French, German, Italian, Portuguese and Spanish. Please consult the Readme file for the latest information. Only one speech system will be installed with OmniPage Pro, depending on your language choice at the start of installation. If you specify a language with no speech system available, English is installed.

If you have other SAPI-compliant speech systems on your computer, they will be detected and available. Their voices will be available in the Speech Settings dialog box. Once you have associated a voice with a language, OmniPage Pro will remember this, and switch voices according to the recognition language of your document.

PAGE OUTLINE

The Page outline window lets you change the order of areas on a page or of paragraphs inside areas. It also lets you define how text should flow if you export with Retain Flowing Columns view. Open the page outline window from the View menu. The areas correspond to the zones used during recognition and also to frames used in the Text Editor. Click and drag an item to the desired location. Reordered paragraphs display immediately in the Text Editor and are exported. Reordered areas display and are exported in No Formatting View and Retain Fonts and Paragraphs view. In True Page view they have no practical effect. In Retain Flowing Columns view, arrows show the order of text flow. Move areas to change this order. The positions of the areas do not change, but the arrows show the changed text flow.

5 Saving and exporting

Once you have acquired at least one image for a document, you can export the image(s) to file. Once you have recognized at least one page, you can export recognition results to a target application by:

1. Saving to file

2. Copying a document to the Clipboard

3. Sending a document as a mail attachment

The document remains in OmniPage SE after export. This allows you to save, copy or send it repeatedly, for example with different formatting levels, using different file types, names or locations. You can also add or rerecognize pages or modify the recognized text.

With automatic processing and using the OCR Wizard, you specify the first saving destination before processing starts. When the last available page is recognized (or proofread, if that was requested), the exporting occurs.

You can specify export any time the program is not busy. If you ask to export a document with unrecognized pages, you will be asked whether they should be recognized first. If you answer *No*, only results from recognized pages will be exported. If zones have been modified on recognized pages, you will be invited to rerecognize those pages before exporting.

PREPARING RECOGNITION RESULTS FOR EXPORT

Text is exported to file, Clipboard or mail with the formatting level defined by the view set in the Text Editor at export time, if that is possible. However, some export file types and target applications cannot support all formatting elements. You may be warned if there is a mismatch and offered the highest permissible view. You can accept that, or cancel export, set a different view and restart the export.

The table in the section *File types for saving recognition results* in chapter 6 tells you which file types support which formatting levels.

Here is how you can use the views for export:

No Formatting view
This view is needed when exporting to ASCII, Unicode or other formats with extension .TXT. These file types cannot accept graphics or tables. Of course, you can export plain text to any file type and target application.

Retain Fonts and Paragraphs view
This is suitable for all formats except those with the TXT or PDF extensions. These formats can all handle graphics and tables.

True Page view
This is suitable only for file types and target applications capable of handling frames or text boxes. When you export to PDF, True Page is used as source, regardless of your editor view (Not applicable to OmniPage SE). The reading order of zones, or areas reordered in the Page outline window have no influence when True Page is used for export.

Retain Flowing Columns view
Set this at export time to keep the original layout of the pages, including columns. This is done wherever possible with column settings, not with frames. Text will then flow from one column to the other, which does not happen when frames are used. Arrows show the text flow order. You can change this order with the Page outline window, as described in *Page outline* in chapter 4.

SAVING TO FILE

You can save recognized pages and original images to disk in a wide variety of file types. See chapter 6 for a complete list of supported file types: *File types for opening and saving images* and *File types for saving recognition results.*

Saving original images

1. Choose Save Image... in the File menu. In the dialog box that appears, select a folder location and a file type for your images. Type in a file name.

2. Select to save the current image only or all images in the document. In the second case you can have all images in a single multi-page image file, providing you set TIFF or DCX as file type. Otherwise each image is placed in a separate file. OmniPage SE adds numerical suffixes to the file name you provide, to generate unique file names.

3. Click OK to save the image(s) as specified. Zones and recognized text are not saved with the file. If possible, the file is saved as displayed: that is black-and-white, grayscale or color. Black-and-white images are saved at their original resolutions. Grayscale and color images are reduced to approximately 150 dpi.

Tip To see the image size and original resolution of an image, hover the cursor over it in the Original Image area or over its thumbnail in the Document Manager.

Note In OmniPage Pro you can save your document to four variants of PDF, including 'image only'. This is saving the recognition results as image, not the original images. PDF saving is not available in OmniPage SE.

Saving recognition results

1. Choose Save As... in the File menu, or click the Export Results button in the Manual OCR toolbar with Save as File selected in the drop-down list.

2. The Save As dialog box appears, as shown in its expanded form.

Select this to automatically open the saved file in its target application.

Select this to have the paragraph styles from the Text Editor exported with the recognized text.

Click Advanced to open the lower panel and Basic to close it.

Choose from:

Create one file for all pages
Create one file per page
Create a new file at each blank page.
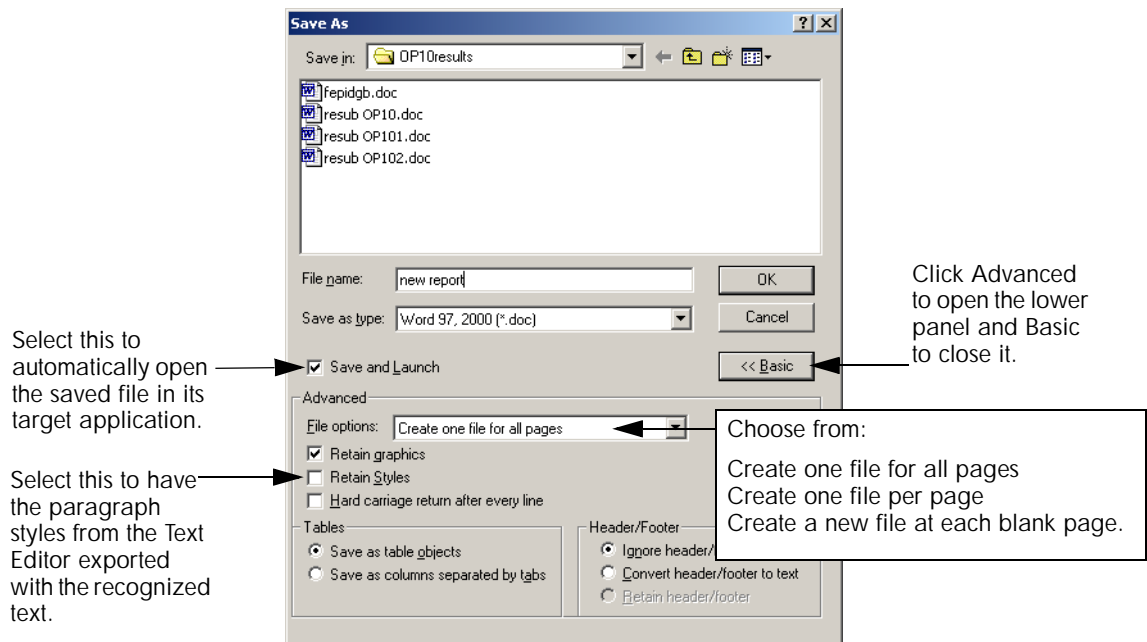
3. Select a folder location and a file type for your document. The special OPD file type is the last in the file type list.

4. Type in a file name. Click the Advanced button to see all the saving options. Select these as desired.

5. Click OK. The document is saved to disk as specified. If 'Save and Launch' is selected, the exported file will appear in its target application; that is the one associated with the selected file type in your Windows system.

Note   Graphics and formatting are saved in the document only if the selected file type supports them. The formatting level for export is the Editor view set at saving time. You will be warned if the formatting level is not supported by the export file type.

Note   If more than one export file is created, OmniPage SE will append a numerical suffix to your file name to create unique file names. If you select 'Create a new file at each blank page' with input from image files, see how to place blank images in the section *Input from image files* in chapter 3.

SAVING A DOCUMENT AS YOU WORK

Click the Save button in the Standard toolbar or choose Save in the File menu to save changes to the current document as you work. If you do this with an untitled document, the Save As dialog box appears.

With a named document, the Save command saves it to the name and format of its last save, as displayed in the title bar. If the document was last saved as an OmniPage Document, the save command updates this document: new or changed images, changed zoning, recognition results and training are all saved. If the document was last saved to a text-based file type, only changes to the recognition results are saved.

If you want to work with your document again in OmniPage SE in a later session, save it as an OmniPage Document. This is a special output file type. It saves the original images together with the recognition results, settings and training. See the section *OmniPage Documents* in chapter 2. The Save As dialog box lists available file types in its Save as Type drop-down list. The OmniPage Document is the last format in the list.

Your OmniPage Documents can be passed between OmniPage SE and OmniPage Pro 11. In OmniPage SE any training data in the OPD is ignored and training cannot be done.

If you first save the document as an OmniPage Document (for instance as memo.opd), then modify it and later save it to a text file (for instance as memo.txt), then modify it again and click Save, the recent changes are saved to the memo.txt file, not to the OPD. When you close the document or exit the program, you will be prompted to save the document if it has not been saved as an OmniPage Document, or there are changes since the last OPD save.

COPYING A DOCUMENT TO THE CLIPBOARD

You can copy the recognition results from every recognized page of a document to the Clipboard. The copying is reported by a progress monitor. You can then paste the Clipboard contents into another application.

Text formatting, such as bold and italics, is retained when you paste into an application that supports RTF information. Otherwise, only plain text will be pasted. Graphics are retained if the application supports insertion of images.

▼ **To copy a document to the Clipboard**
- With automatic processing, select Copy to Clipboard as the command in the Export Results drop-down list on the AutoOCR toolbar or in the OCR Wizard. The text is sent to Clipboard as soon as the last available page is recognized or proofed.
- With manual processing, select the Copy to Clipboard command in the Export Results drop-down list and then click its button on the Manual OCR toolbar. Copying starts immediately.

SENDING A DOCUMENT AS A MAIL ATTACHMENT

You can send recognition results as one or more files attached to a mail message if you have installed a MAPI-compliant mail application, such as Microsoft Outlook.

▼ **To send a document by e-mail**

- With automatic processing, select Send as Mail as the command in the Export Results drop-down list on the AutoOCR toolbar. The Send Mail dialog box appears as soon as the last available page in the document is recognized or proofed.

- With manual processing, select Send as Mail as the command in the Export Results drop-down list and then click its button on the Manual OCR toolbar. The Send Mail dialog box appears immediately.

At any time the program is not busy, choose Send as Mail in the File menu to call up the Send Mail dialog box.
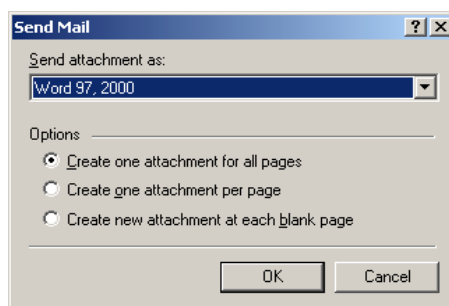
1. The Send Mail dialog box lets you specify a file type and attachment options: one attachment for all pages, one attachment per page, new attachment at each blank page. Set all options and click OK.

2. Log into your mail application if you are prompted to do so.

3. Your mail application appears with the attachment(s) in a new empty message. Attachments take the name used for the last save of the document in OmniPage SE, or 'Untitled from OmniPage'. The suitable file extension is added, and numerical suffixes for multiple attachments.

4. Address your mail message, add message text as desired and click the Send button.

6 Technical information

This chapter provides troubleshooting and other technical information about using OmniPage SE.

Please also read the online Readme file and other help topics, or visit the ScanSoft web pages. The Scanner Information web page contains detailed and regularly updated information about scanner setup and support. The Readme file contains last-minute information relating to OmniPage SE. Access to the Readme file and to ScanSoft's web pages is provided in the Help menu.

This chapter contains the following information:

◆ Troubleshooting
  • Solutions to try first
  • Testing OmniPage SE
  • Low memory problems
  • Low disk space problems
◆ Supported file types
  • File types for opening and saving images
  • File types for saving recognition results
  • Saving to PDF
◆ OCR problems
  • Text does not get recognized properly
  • Problems with fax recognition
  • System or performance problems during OCR
◆ Uninstalling the software

TROUBLESHOOTING

Although OmniPage SE is designed to be easy to use, problems sometimes occur. Many of the error messages contain self-explanatory descriptions of what to do – check connections, close other applications to free up memory, and so on. Sometimes that is all the troubleshooting help you need.

Please see your Windows documentation for information on optimizing your system and application performance.

Solutions to try first

Try these solutions if you experience problems starting or using OmniPage SE:

◆ Make sure that your system meets all requirements listed under *System requirements* in chapter 1.

◆ Make sure that your scanner is plugged in and that all cable connections are secure.

◆ Visit the support section of ScanSoft's web site at www.scansoft.com. It contains Tech Notes on commonly reported issues using OmniPage SE. Our web pages may also offer assistance on the installation process and troubleshooting.

◆ Turn off your computer and your scanner, turn your scanner back on, and then restart your computer. Make sure other applications are functioning properly.

◆ Use the software that came with your scanner to verify that the scanner works properly before using it with OmniPage SE.

◆ Make sure you have the correct drivers for your scanner, printer, and video card. Visit the ScanSoft's Scanner Information web page through the Help menu for more information.

◆ Run ScanDisk for Windows 95, 98 or Me, or Check Disk for Windows NT and Windows 2000 to check your hard disk for errors. See Windows online Help for more information.

◆ Defragment your hard disk. See Windows online Help for more information.

◆ Uninstall and reinstall OmniPage SE, as described in *Uninstalling the software* at the end of this chapter.

Testing OmniPage SE

Restarting Windows 95, 98, 2000 or Me in safe mode or Windows NT in VGA mode allows you to test OmniPage SE on a simplified system. This is recommended when you cannot resolve crashing problems or if OmniPage SE has stopped running altogether. See Windows online Help for more information.

Note Your scanner will not run with OmniPage SE in safe mode or VGA mode, so do not test scanner problems in this configuration.

▼ **To test OmniPage SE in safe mode (Windows 95, 98, 2000 or Me):**

1. Restart your computer in safe mode by pressing F8 immediately after you see the 'Starting Windows' message.

2. Launch OmniPage SE and try performing OCR on an image. Use a known image file, for instance one of the supplied sample image files.

   - If OmniPage SE does not launch or run properly in safe mode, then there may be a problem with the installation. Uninstall and reinstall OmniPage SE (see *Uninstalling the software*), and then run it in Windows safe mode.
   - If OmniPage SE runs in safe mode, then a device driver on your system may be interfering with OmniPage SE operation. Troubleshoot the problem by restarting Windows in Step-by-Step Confirmation mode. See Windows online Help for more information.

▼ **To test OmniPage SE in VGA mode (Windows NT):**

1. Restart your computer.

2. Select Windows NT Workstation Version 4.00 [VGA mode] and press Enter.

3. Press Ctrl+Alt+Del and select Task Manager.

4. In the Task Manager dialog box, select all background applications and click End Process. See Windows online Help for more information.

5. Launch OmniPage SE and try performing OCR on an image. Use a known image file such as one of the supplied sample files.

Note You can also run OmniPage SE from a command line in its own safe mode. Choose Start ➤ Run, browse for the file OmniPage.exe and add the command line option /safe. This starts the program, but ignores previously stored settings and does not try to recover a document from an abnormal termination.

Low memory problems

OmniPage SE may run poorly under low-memory conditions. This may be indicated by various error messages or if OmniPage SE works slowly and accesses the hard drive often. Try these solutions for low memory conditions:

◆ Restart your computer.
◆ Close other open applications to release memory.
◆ Close unnecessary OmniPage SE applications.
◆ Defragment your hard disk to free up contiguous blocks of disk space. See Windows online Help for instructions.
◆ Increase the amount of free hard disk space.
◆ Increase your computer's physical memory (RAM).
◆ More memory optimizes OCR performance. See *System requirements* in chapter 1 for more information.

Low disk space problems

Problems may occur if your system runs low on free disk space. Try these solutions for low disk space problems:

◆ Empty the Windows Recycle Bin.
◆ Close all open applications and delete the *.tmp files in the Temp folder. This folder is usually located in your Windows folder.
◆ Run ScanDisk for Windows 95, 98 or Me, or Check Disk for Windows NT or Windows 2000 to check for errors that may be using disk space. See Windows online Help for instructions.
◆ Back up unneeded files onto floppy disks or other media and delete them from your hard disk.

◆ Remove Windows applications that you do not use.
◆ Defragment your hard disk. See Windows online Help for instructions.
◆ Clear the cache for your web browser and limit its size.

SUPPORTED FILE TYPES

The program supports a wide range of file types. Several important types have been added in OmniPage SE.

File types for opening and saving images

File type	Extension	Multi-page	Open / Save	B/W, Grayscale, Color
BMP, Bitmap	*.bmp	No	Open and Save	All
DCX	*.dcx	Yes	Open and Save	All
GIF	*.gif	N/A	N/A	N/A
JPEG	*.jpg	No	Open and Save	Grayscale, color
PCX	*.pcx	No	Open and Save	All
PDF	*.pdf	N/A	N/A *(see note)*	N/A
PNG	*.png	No	Open and Save	All
TIFF Compressed G3	*.tif	Yes	Open	B/W
TIFF Compressed G4	*.tif	Yes	Open and Save	B/W
TIFF Compressed LZW	*.tif	N/A	N/A	N/A
TIFF FX	*.xif	N/A	N/A	N/A
TIFF PackBits	*.tif	Yes	Open and Save	All
TIFF Uncompressed	*.tif	Yes	Open and Save	All

Input image files can have resolutions up to 600 dpi, but 300 dpi (both horizontally and vertically) is recommended for optimum OCR accuracy. The program stores black-and-white images at their original resolution, but grayscale and color images are not usually saved above 150 dpi. Hover the cursor over an image for a popup window showing the size and resolution of the original image.

Note If you try to save a black-and-white image to JPEG format, the program will offer conversion to grayscale. With TIFF G3 and G4 it will offer conversion to black-and-white.

Note Saving to PDF format is supported in OmniPage Pro 11, with four options. One of these is to export image only. But this exports the recognition results as images, not the original images, through the Save As dialog box. This is not available in OmniPage SE. Also, OmniPage SE cannot handle GIF, LZW TIFF and TIFF FX files.

File types for saving recognition results

File type	Exten-sion	Format levels (Text Editor views)	Supports graphics
ASCII text [1]	*.txt/.csv	No Formatting view (NFV)	No
Adobe PDF, normal	*.pdf	N/A	N/A
Adobe PDF with image substitutes	*.pdf	N/A	N/A
Adobe PDF with image on text	*.pdf	N/A	N/A
Adobe PDF, image only	*.pdf	N/A	N/A
Excel (3.0 to 7.0, 97, 2000)	*.xls	NFV, RFP (Spreadsheet)	Yes
FrameMaker (5.5.3)	*.mif	All	Yes
Freelance Graphics	*.txt	No Formatting view (NFV)	No
Harvard Graphics	*.txt	No Formatting view (NFV)	No
HTML (3.2 or 4.0) [2]	*.htm	All	Yes [2]
PowerPoint 97	*.rtf	All	Yes
Microsoft Publisher 98	*.rtf	All	Yes
Word for Windows (6.0, 97, 2000)	*.doc	All	Yes
PageMaker 6.5.2	*.doc	All	Yes
Quattro Pro for Windows 4.0, 8	*.xls	NFV, RFP (Spreadsheet)	No
Rich Text Format (RTF) 6.0/95 [3]	*.rtf	All	Yes
Unicode Text [1][4]	*.txt/.csv	No Formatting view (NFV)	No
Ventura Publisher	*.doc	All	Yes
WordPad	*.rtf	NFV, RFP [5]	Yes
WordPerfect (5.1,5.2,6.0,6.1,8,9,10)	*.wpd	All	Yes
OmniPage Document [6]	*.opd	All	Yes

1   ASCII and Unicode text can be with flowing text, with line breaks or comma separated. The latter have the extension .csv and are used for plain text input of tables into spreadsheet programs.

2   When saving to HTML, all graphics are saved as separate image files using JPEG format. HTML 4.0 is supported only in OmniPage Pro 11, OmniPage SE support is limited to HTML 3.2.

3   Recognition results are sent to Clipboard in this format and will be pasted in RTF if possible, and as Unicode or ASCII text if not.

4   Unicode text can handle the widest range of accented characters.

5   True Page or Retain Flowing Columns (RFC) views will not be refused, but will appear as Retain Fonts and Paragraphs (RFP) view, that is, without columns.

6   OmniPage Documents created by OmniPage SE or OmniPage Pro 11 can be reopened by OmniPage SE. It can also open OPD files created by OmniPage Pro 10 and the similar MET files from OmniPage Pro 9. These files remain in their old format and a copy is converted to OmniPage SE.

Saving to PDF

This section does not apply to OmniPage SE.

In OmniPage Pro 11, you have four choices when saving recognition results to Portable Document Format (PDF) files.

Normal:
Pages are exported as they appeared in the Text Editor in True Page view. The PDF file can be viewed and searched in a PDF viewer and edited in a PDF editor.

With image substitutes:
As above, but reject and suspect characters have image overlays, so these uncertain characters display as they were in the original document. The PDF file can be viewed, searched and edited.

Image only:
The PDF file is viewable only and cannot be modified in a PDF editor and text cannot be searched.

Image on text:
The PDF file is viewable only and cannot be modified in a PDF editor. But there is a linked text file behind each image, so the text can be searched. A found word is highlighted in the image.

OCR PROBLEMS

This section contains information and solutions for possible OCR problems. First we provide suggestions for improving recognition accuracy, second on getting good results from fax input and finally on system or performance problems arising during OCR.

Text does not get recognized properly

Try these solutions if any part of the original document is not converted to text properly during OCR:

◆ Look at the original page image and ensure that all text areas are enclosed by text zones. If an area is not enclosed by a zone, it is generally ignored during OCR. See *Manual zoning* in chapter 3.

◆ Make sure text zones are identified correctly. Reidentify zone types and contents, if necessary, and perform OCR on the document again. See *Zone properties* in chapter 3.

◆ Be sure you do not have an unsuitable template loaded by mistake. If zone borders cut through text, recognition is impaired.

◆ Adjust the brightness and contrast sliders in the Scanner panel of the Options dialog box. You may need to experiment with different settings combinations to get the desired results.

◆ Check the resolution of the original image. Hover the cursor over the Original Image area for a popup display. If the resolution is significantly above or below 300 dpi, recognition is likely to suffer.

◆ Make sure the correct document languages are selected in the OCR panel of the Options dialog box. Only languages included in the document should be selected.

◆ Turn IntelliTrain on and make some proofing corrections. This is most likely to help with stylized fonts or uniformly degraded documents. If IntelliTrain was running, try turning it off – on some types of degraded documents it may not be able to help. This does not apply to OmniPage SE.

◆ If you use True Page as the Text Editor view or for export, recognized text is put into frames (formatting boxes). Some text may be hidden if a frame is too small. To view the text, place the cursor in the text frame and use the arrow keys on your keyboard to scroll to the top, bottom, left, or right of the frame.

◆ Check the glass, mirrors, and lenses on your scanner for dust, smudges, or scratches. Clean if necessary.

Note   OmniPage SE only recognizes machine printed-text characters such as typewritten or laser-printed text. It can handle dot-matrix characters, though accuracy may be lower on draft-quality texts. It cannot read handprint or handwriting. However, it can retain signatures or other handwritten text as a graphic.

Problems with fax recognition

Try these solutions to improve OCR accuracy on fax images:

◆ Ask senders to use clean, original documents if possible.

◆ Ask senders to select Fine or Best mode when they send you a fax. This produces a resolution of 200 x 200 dpi.

◆ Ask senders to transmit files directly to your computer via fax modem if you both have one. You can save fax images as image files and then load them into OmniPage SE. See the section *Input from image files* in chapter 3.

System or performance problems during OCR

Try these solutions if a crash occurs during OCR or if processing takes a very long time:

◆ Resolve low memory problems. See *Testing OmniPage SE*.

◆ Resolve low disk space problems. See *Testing OmniPage SE*.

◆ Minimize all applications or click Alt+Tab to check for Windows error messages.

◆ Check the quality of the image you are recognizing.

◆ Consult your scanner documentation on ways to improve the quality of scanned images.

◆ Break complex page images (lots of text and graphics or elaborate formatting) into smaller jobs. Draw zones manually or modify automatically created zones and perform OCR on one page area at a time. See *Working with zones* in chapter 3 on creating and modifying zones.

◆ Restart Windows 95, 98 and Me and 2000 in safe mode, or Windows NT and in VGA mode and test OmniPage SE by performing OCR on the included sample image files. See the section *Testing OmniPage SE*.

If you are performing multiple tasks at once, such as recognizing and printing, OCR may take longer.

UNINSTALLING THE SOFTWARE

Sometimes uninstalling and then reinstalling OmniPage SE will solve a problem. You should uninstall OmniPage SE before installing OmniPage Pro 11 or any OmniPage evaluation software. OmniPage SE's Uninstall program will not remove any of the following user-created files:

Zone templates (*.zon)

Training files (*.otd) (Not applicable to OmniPage SE)

User dictionaries (*.ud)

OmniPage Documents (*.opd)

To uninstall from Windows NT or Windows 2000, you must be logged into your computer with administrator privileges.

▼ **To uninstall or reinstall OmniPage SE:**

◆ Close OmniPage SE.

◆ Click Start in the Windows taskbar and choose Settings ➤ Control Panel ➤ Add/Remove Programs.

◆ Select OmniPage SE and click Change.

◆ Click Next in the dialog box that appears.

◆ Select Remove or Repair, then Next.

◆ Follow instructions until the process is finished.

I N D E X