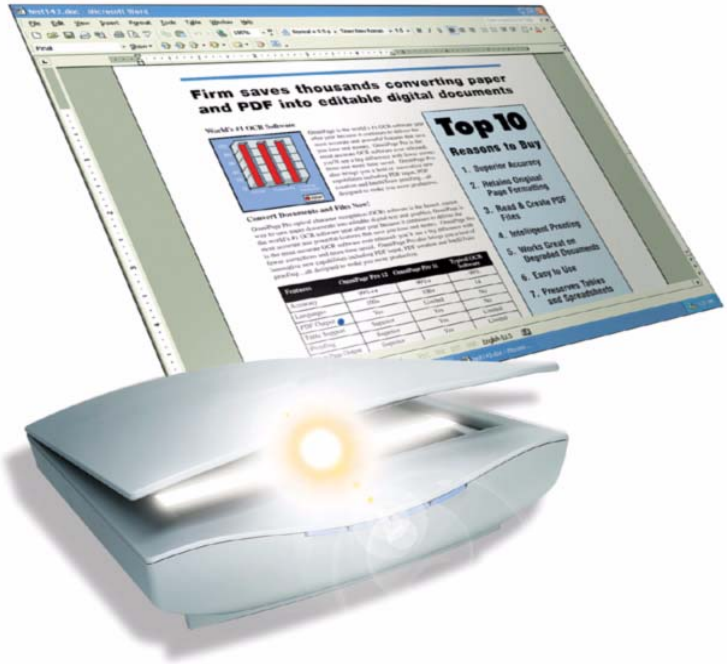
ScanSoft®



OmniPage® PRO 12

User's Guide

Legal Notices

Copyright © 2002 ScanSoft, Inc. All rights reserved. No part of this publication may be transmitted, transcribed, reproduced, stored in any retrieval system or translated into any language or computer language in any form or by any means, mechanical, electronic, magnetic, optical, chemical, manual, or otherwise, without prior written consent from ScanSoft, Inc., 9 Centennial Drive, Peabody, Massachusetts 01960. Printed in the United States of America and in the Netherlands.

The software described in this book is furnished under license and may be used or copied only in accordance with the terms of such license.

Important Notice

ScanSoft, Inc. provides this publication "as is" without warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability or fitness for a particular purpose. Some states or jurisdictions do not allow disclaimer of express or implied warranties in certain transactions; therefore, this statement may not apply to you. ScanSoft reserves the right to revise this publication and to make changes from time to time in the content hereof without obligation of ScanSoft to notify any person of such revision or changes.

Trademarks and Credits

ScanSoft, OmniPage, OmniPage Pro, PaperPort, *Pagis, True Page* and *Direct OCR* are registered trademarks or trademarks of ScanSoft, Inc., in the United States and/or other countries.

All other company names or product names referenced herein may be the trademarks of their respective holders.

ScanSoft, Inc.
9 Centennial Drive
Peabody, MA 01960
U.S.A.

ScanSoft Belgium BVBA
Guldensporenpark 32
BE-9820 Merelbeke
Belgium

Part Number 50-281201-00A

CONTENTS

WELCOME 7

Using this Guide 8

Getting online Help 9

Online HTML Help 9

Context-Sensitive Help 9

Tech Notes 10

Glossary 10

1 INSTALLATION AND SETUP 11

System requirements 12

Installing OmniPage Pro 13

Setting up your scanner with OmniPage Pro 14

How to start the program 16

Registering your software 17

New features in OmniPage Pro 12 17

2 INTRODUCTION 19

What is optical character recognition 20

OmniPage Pro's OCR capabilities 20

Documents in OmniPage Pro 21

Basic processing steps 21

The OmniPage Desktop 22

The Menu bar 23

The Toolbars 23

The Image Panel 24

The Text Editor 24
The OmniPage Toolbox 25
Managing documents 26
Thumbnails 26
Document Manager 27
Customizing Document Manager columns 28
Deleting pages from a document 28
Printing a document 29
Closing a document 29
OmniPage Documents 29
Why save to OPD 30
How to save to OPD 30
Settings 31

3 PROCESSING DOCUMENTS 33

Quick Start Guide 34
Loading and recognizing sample image files 34
Scanning and recognizing a single page 34
Processing overview 36
Automatic processing 38
Stopping and restarting automatic processing 39
Manual processing 40
Combined processing 41
Processing with the OCR Wizard 43
Processing from other applications 45
How to set up Direct OCR 45
How to use Direct OCR 45
How to use OmniPage Pro with PaperPort 46
Processing with Schedule OCR 47
Defining the source of page images 48
Input from image files 48

Input from scanner 49
Scanning with an ADF 50
Scanning without an ADF 51
Describing the layout of the document 51
Zones and backgrounds 53
Automatic zoning 53
Manual zoning 54
Zone types and properties 55
Working with zones 57
Table grids in the image 59
Using zone templates 61

4 PROOFING AND EDITING 63

The editor display and views 64
Proofreading OCR results 65
Verifying text 66
User dictionaries 68
Training 69
Manual training 69
IntelliTrain 70
Training files 71
Text and image editing 72
On-the-fly editing 74
Reading text aloud 75

5 SAVING AND EXPORTING 77
Saving original images 78
Saving recognition results 79
Saving a document as you work 80
Selecting a formatting level 81
Selecting advanced saving options 82
Saving to PDF 84
Copying pages to Clipboard 84
Sending pages by mail 85

6 TECHNICAL INFORMATION 87
Troubleshooting 88
Solutions to try first 88
Testing OmniPage Pro 89
Increasing memory resources 90
Increasing disk space 90
Text does not get recognized properly 91
Problems with fax recognition 92
System or performance problems during OCR 92
ODMA support 93
Advanced features in Schedule OCR 93
Supported file types 94
File types for opening and saving images 94
File types for saving recognition results 95
Uninstalling the software 96

INDEX 97

Welcome

Welcome to OmniPage Pro®, 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 Pro 12. 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 Pro'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. See "Getting online Help" on page 9.

Readme File

The Readme file contains last-minute information about the software. Please read it before using OmniPage Pro. To open this HTML file, choose Readme in the OmniPage Pro 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; ScanSoft tests the 25 most widely used scanner models. Access ScanSoft's web site from the OmniPage Pro 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 Pro 12. 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 topics in the online Help system. Presents longer option texts in dialog boxes.
Non-serif	Presents file names: sample.tif
	A note presents an item of additional information.
	A tip presents ideas for using program features to accomplish specific tasks.

Getting online Help

In addition to using this guide, you can use OmniPage Pro's online Help to learn about features, settings, and procedures. Online Help is available after you install OmniPage Pro.

Online HTML Help

Open OmniPage Pro's online Help at its top level by choosing 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 Pro menu item, toolbar button, screen area or dialog box, in the following ways:

Click the Help tool 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. Use Shift + F1 to get context-sensitive help for shortcut menu items.

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 Pro 12. 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.

Chapter 1

Installation and setup

This chapter provides information on installing and starting OmniPage Pro 12. It presents the following topics:

- System requirements
- Installing OmniPage Pro
- Setting up your scanner with OmniPage Pro
- How to start the program
- Registering your software
- New features in OmniPage Pro 12

System requirements

You need the following minimum system requirements to install and run OmniPage Pro 12:

- A computer with a Pentium or higher processor
- Microsoft Windows 98 (from second edition), Windows Me, Windows NT 4.0 (with at least Service Pack 6), Windows 2000 or Windows XP
- 64MB of memory (RAM), 128MB recommended
- 90MB of free hard disk space for the application files plus 5MB working space during installation
- 5MB for Microsoft Installer (MSI) if not present (This is present as part of the operating system in Windows Me, Windows 2000 and Windows XP)
- SVGA monitor with 256 colors, but preferably 16-bit color (called High Color in Windows 2000 and Medium Color in XP) and 800 x 600 pixel resolution
- Windows-compatible pointing device
- CD-ROM drive for installation
- A compatible scanner with its own scanner driver software, 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.

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

Installing OmniPage Pro

OmniPage Pro 12's installation program takes you through installation with instructions on every screen.

Before installing OmniPage Pro:

- Close all other applications, especially anti-virus programs.
- Log into your computer with administrator privileges if you are installing on Windows NT, 2000 or XP.
- If you own a previous version of OmniPage Pro, or if you are upgrading from demonstration software or an OmniPage Special Edition, the installer asks your consent to uninstall that product.

▼ **To install OmniPage Pro:**

1. Insert OmniPage Pro'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 Pro 12, but your choice at installation time determines which Text-to-Speech system will be installed with the program. See the second note below.

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

Sometimes uninstalling and then reinstalling OmniPage Pro will solve a problem. See "Uninstalling the software" on page 96.

It is planned to provide Text-to-Speech for English, French, German, Italian, Portuguese and Spanish. This may vary depending on region or version. The Readme file provides latest information. A speech system for only one language can be installed with OmniPage Pro. See "Reading text aloud" on page 75.

Setting up your scanner with OmniPage Pro

All files needed for scanner setup and support are copied automatically during the program's installation. Before using OmniPage Pro 12 for scanning, your scanner should be installed with its own scanner driver software and tested for correct functionality. Scanner driver software is not included with OmniPage Pro.

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.

Please follow these steps to use the Scanner Wizard to setup your scanner with OmniPage Pro 12:

- Choose Start ▸ Programs ▸ ScanSoft OmniPage Pro 12.0 ▸ Scanner Wizard

  or click the Setup button in the Scanner panel of the Options dialog box.

  or choose a scan setting in the Get Page drop-down list in the OmniPage Toolbox and click the Get Page button.

  The Scanner Setup Wizard starts. The first panel appears only on first setup when called from inside OmniPage Pro.
- Choose 'Select scanner or digital camera', then click Next. You see a list of all detected TWAIN scanner drivers, with the system default scanner selected.
- Click once to select the driver of the scanner you want to use. Click 'Other drivers...' if you need to browse for a driver. Select 'Configure Advanced Settings' for an extra panel if you want your scanner's own interface to be hidden during scanning or to modify the image transfer method. Click on Next.
- Choose Yes to test your scanner configuration, then click Next. The wizard will now test the connection from the computer to your scanner. When completed, 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 your scanner to support 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 Pro 12! Click on Finish.

To change the scanner settings at a later time, or to set up a different scanner, reopen the Scanner Setup Wizard from the Windows Start menu or from the Scanner panel of the Options dialog box. To test and repair an improperly functioning scanner, open the Scanner Setup Wizard from the Windows Start menu and select ‘Test scanner or digital camera’ in the first panel, then work through the procedure described above.

How to start the program

To start OmniPage Pro 12 do one of the following:

- Click Start in the Windows taskbar and choose Programs ▸ ScanSoft OmniPage Pro 12.0 ▸ OmniPage Pro 12.0.

- Double-click the OmniPage Pro 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" on page 29.

On opening, OmniPage Pro's title screen is displayed and then its desktop. See "The OmniPage Desktop" on page 22. It provides an introduction to the program's main working areas.

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 Pro 12.0 ▸ Schedule OCR. See "Processing with Schedule OCR" on page 47.
- Click Acquire Text from the File menu of an application registered with the Direct OCR™ facility. See "How to set up Direct OCR" on page 45.
- 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 Pro 12 with ScanSoft's PaperPort® or Pagis® document management products, to add OCR services. See "How to use OmniPage Pro with PaperPort" on page 46.

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 12

The OmniPage® product family is augmented by OmniPage Pro 12. If you are upgrading, you may not need to consult this guide very much. Here are some main areas of innovation compared to OmniPage Pro 11:

- **Dramatic increase in accuracy**
  Improved synergy between recognition engines, support for professional dictionaries and the ability to train characters chosen by the user boost accuracy to new levels.
- **Streamlined interface**
  Automatic and manual processing are now driven directly from the OmniPage Toolbox without separate toolbars. See page 25. Thumbnails now display in the Image Panel; choose to see the current page, thumbnails or both. See page 26. The previous Detail view becomes the Document Manager and includes a Note column for comments and searchable keywords.
- **New zoning concepts**
  On-the-fly zoning allows zone changes to be processed immediately without having to re-recognize the whole page. See page 74. Page backgrounds are defined as process (auto-zone) or ignore, so all zoning instructions appear on the page and can be

saved to zone templates. See page 53. Irregular zones can be drawn and zones split and joined more simply, without the need for separate tools. See page 57.

- **Better proofing and verifying**
  The Proofing dialog box now shows suspect words in a wider context. A dynamic verifier can stay open as text is being checked, with the image display and window tracking the editing position. See page 65.

- **Formatting levels for display and saving**
  There are three formatting levels for Text Editor display. See page 64. The output formatting level is now chosen at export time; the choices depend on the specified file type. An export choice 'Flowing Page' is an improved version of the previous 'Retain Flowing Columns' view. It preserves page layout without boxes and frames whenever possible, so text can flow between columns. See page 81.

- **Superior page analysis**
  The transfer of table formatting has improved, in particular the detection of tables without gridlines in original pages. Web and e-mail addresses can be detected and transferred to the Text Editor; hyperlinks can be inserted. Reading order can now be viewed and changed after recognition in the Text Editor's True Page® view. See from page 72.

- **Improved PDF handling**
  OmniPage Pro 12 searches background text in PDFs it opens, to deliver higher recognition accuracy. A new file type 'PDF Edited' allows good format retention on pages that were modified in the Text Editor after recognition.

- **Advanced saving options**
  A wider range of saving options is offered for each output file type. User-defined output file types can be created with customized settings. See page 82. If your edition of OmniPage Pro 12 includes the new saving formats XML and eBook, see page 95.

Chapter 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.

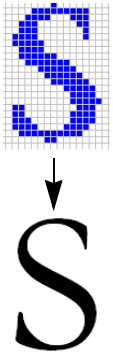
For example, if you want to incorporate information from a magazine article in a report you are preparing, you somehow have to get the text from the article into your computer. Painstakingly retyping the article is not an appealing solution.

This chapter introduces you to the solution: optical character recognition (OCR). It describes how OmniPage Pro 12 uses OCR technology to transform text from scanned pages or image files into editable text for use in your favorite computer applications.

We present the following topics:

- What is optical character recognition
  - Documents in OmniPage Pro
  - Basic processing steps
- The OmniPage 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 Pro 12 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 Pro's OCR capabilities

In addition to text recognition, OmniPage Pro 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 Pro retains are determined by the settings you select. Refer to the *Settings Guidelines* in the online Help for more information about selecting settings.

OmniPage Pro 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 Pro

OmniPage Pro 12 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 Pro 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. See "The OmniPage Desktop" on page 22.

Basic processing steps

There are two main ways of handling documents: with automatic processing or manual processing. See "Automatic processing" on page 38 and "Manual processing" on page 40. The basic steps for both processing methods are broadly the same:

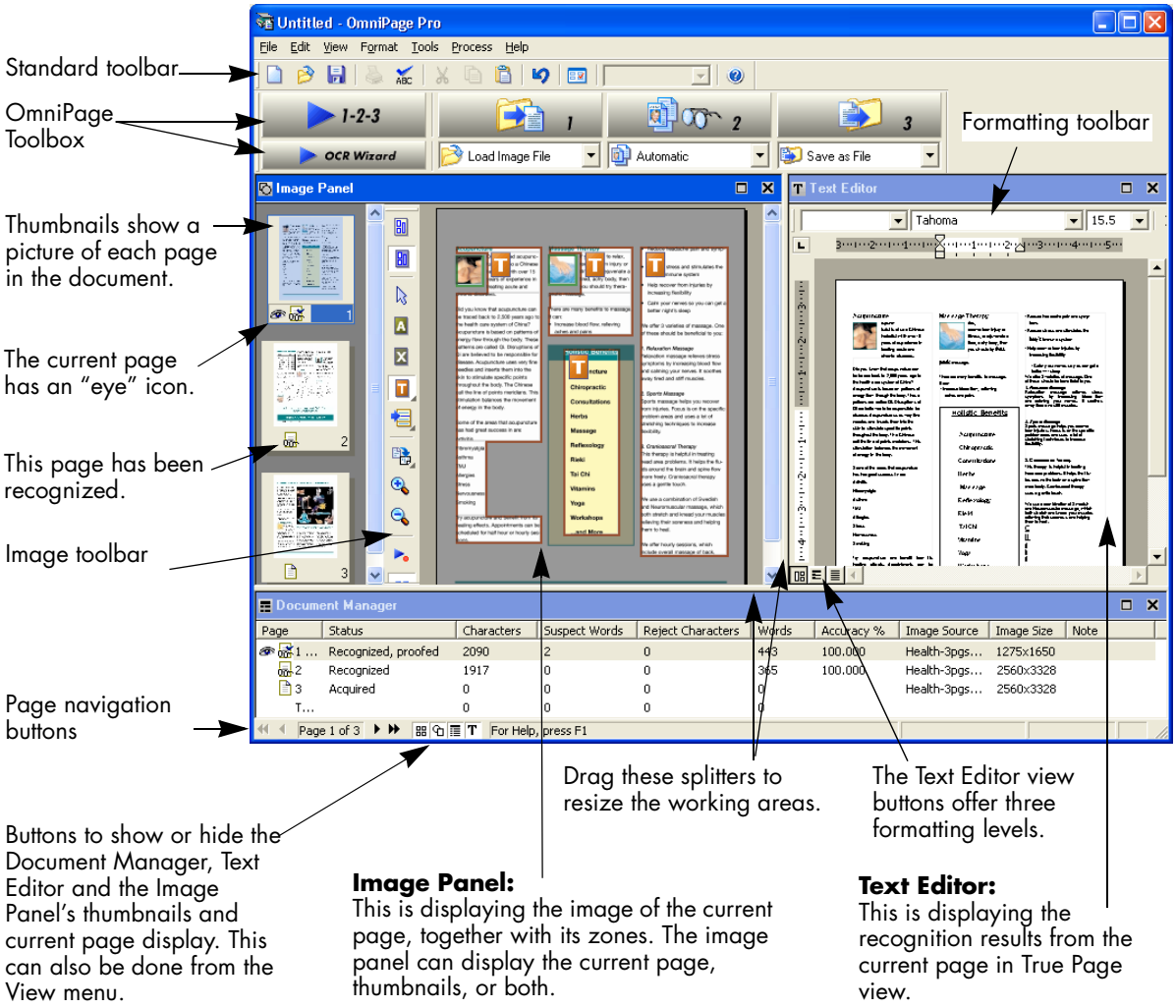
1. **Bring a set of images into OmniPage Pro.**
   You can scan a paper document with or without an Automatic Document Feeder (ADF) or load one or more image files. The resulting images can appear as thumbnails in the Image Panel along with the image of the first page entered. The document pages are summarized in the Document Manager. See "Defining the source of page images" on page 48.

2. **Perform OCR to generate editable text.**
   During OCR, OmniPage Pro 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 "Saving and exporting" on page 77.

The OmniPage Desktop

The OmniPage 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 Image Panel and the Text Editor. Each has close, maximize and restore buttons top right. The Image Panel has an Image toolbar and the Text Editor has a Formatting toolbar.



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 of the thumbnail images show page status.

Status bar buttons let you show or hide 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 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 Toolbars

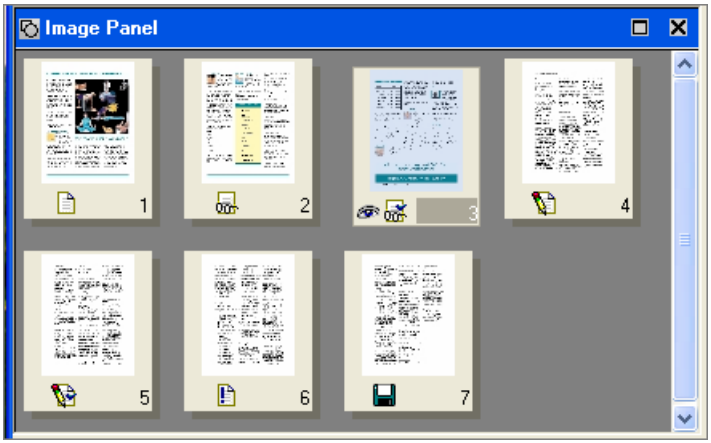
The program has three main toolbars; all can be floated. Use the View menu to show, hide or customize them. Context-sensitive help explains the purpose of all tools. Two further toolbars govern specific tasks.

Toolbar	Default location	Other docking locations	Purpose
Standard	Horizontal under Menu bar	Any edge of the OmniPage Desktop	Performing basic program functions. See page 29 and page 65.
Image	Vertically to left of current page image	Vertically to right of current page image	Image, zoning and table operations. See page 53 and page 59.
Formatting	Horizontal at top of Text Editor	None	Formatting recognized text in the Text Editor. See page 72.
Verifier	Hover the cursor over the verifier window to see this floating toolbar.		Controlling the location and appearance of the verifier. See page 66.
Reorder	Click the Change reading order tool. This toolbar replaces the Formatting toolbar.		Modifying the order of elements in recognized pages. See page 72.

The Image Panel

When this displays the current page image, the Image toolbar is available. All page images have a background value: process or ignore. Zones can be manually drawn on page images, or can be placed automatically after recognition. There are five zone types: Process, Ignore, Text, Table, Graphics. Areas inside process zones and on a process background outside other zones have zones automatically drawn and their zone types determined during processing. See “Zones and backgrounds” on page 53.

If the current page image is hidden, the thumbnails appear in rows to make the best use of the available space.



The Text Editor

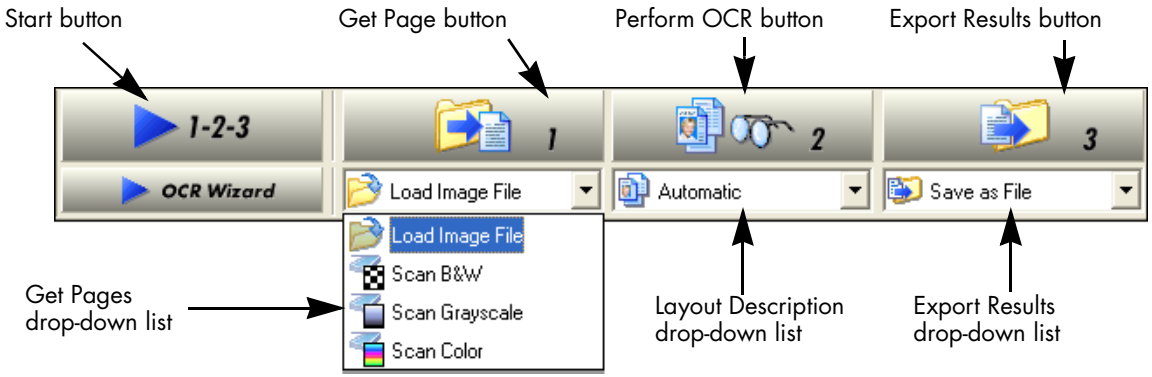
This displays recognition results in any of three formatting levels:

- No Formatting view (NF)
- Retain Fonts and Paragraphs view (RFP)
- True Page (TP)

True Page retains page layout using text, table and picture boxes, and frames. It can display multicolumn areas, to show text blocks that can be treated as flowing columns at export time.
True Page is also an export formatting level, along with **Flowing Page** that retains page layout without boxes and frames. See “The editor display and views” on page 64.

The OmniPage Toolbox

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



Automatic processing is started, and can be stopped and re-started with the Start (1-2-3) button. See "Automatic processing" on page 38.

Manual processing allows you to process documents page-by-page and step-by-step. Start each step with the three large buttons: the Get Page button (1), the Perform OCR button (2) and the Export Results button (3). See "Manual processing" on page 40.

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. See "Combined processing" on page 41.

The **OCR Wizard** is designed for new users. See "Processing with the OCR Wizard" on page 43. If you have a document open when you start the OCR Wizard, the document will be closed after a prompting to save it. 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, re-recognizing pages etc.) with either manual or automatic processing.

Managing documents

Document management can be done by thumbnails in the Image Panel or by the Document Manager, situated along the bottom of the OmniPage Desktop. Both summarize the pages in the document and are synchronized. Our pictures show these with the same seven-page document. Pages 1 and 2 are selected and page 4 is the current page, that is, the one shown in the Image Panel. Page status is shown as follows:

Page	Status	Icon	Page image has been...
1	Acquired		acquired but has not yet been recognized.
2	Recognized		recognized, but not proofread, or proofing was interrupted on the page.
3	Recognized, Proofed		recognized, and proofing has reached the end of the page.
4	Modified		recognized with at least one editing or formatting change made in the Text Editor.
5	Modified, proofed		recognized, edited in the Text Editor, and proofing has reached the end of the page.
6	Pending		acquired, maybe recognized; some zone changes are stored but not yet processed.
7	Saved		recognized and saved at least once.

Thumbnails

These present a set of numbered thumbnail images, one for each page in the document. Scroll to see pages as necessary. The current page has an 'eye' icon. You can select multiple pages in the document; these have a distinctive appearance. Use thumbnails for page operations, as follows:

Jump to a page: Click the thumbnail 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.

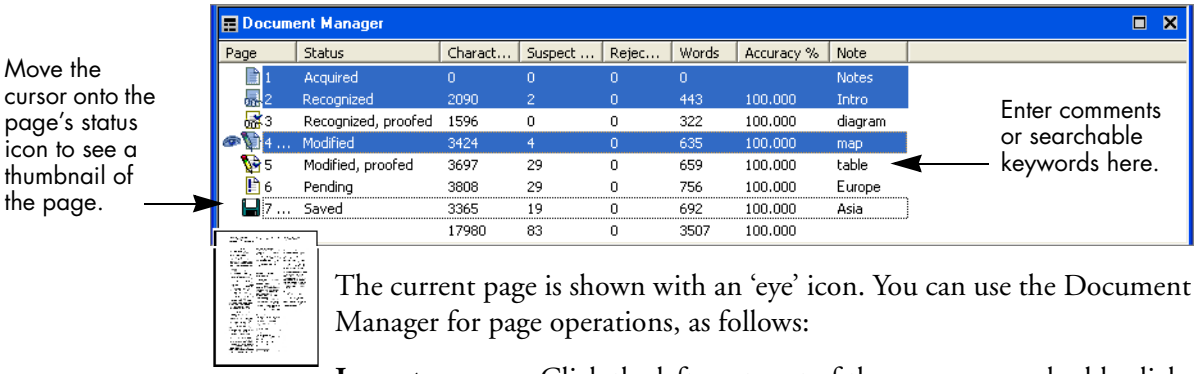
Select multiple pages: Hold down the Shift key and click two thumbnails to select all pages between and including them. Hold down

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. You can also export selected pages.

> Get information on an input image by hovering the cursor over its thumbnail (so long as ToolTips are enabled). A popup text displays the image size in pixels and the program's unit of measurement. Image resolution is also shown.

Document Manager

This 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 shows columns that a user has specified.



The current page is shown with an 'eye' icon. You can use the Document Manager for page operations, as follows:

Jump to a page: Click the leftmost part of the page row or double click anywhere in its row.

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

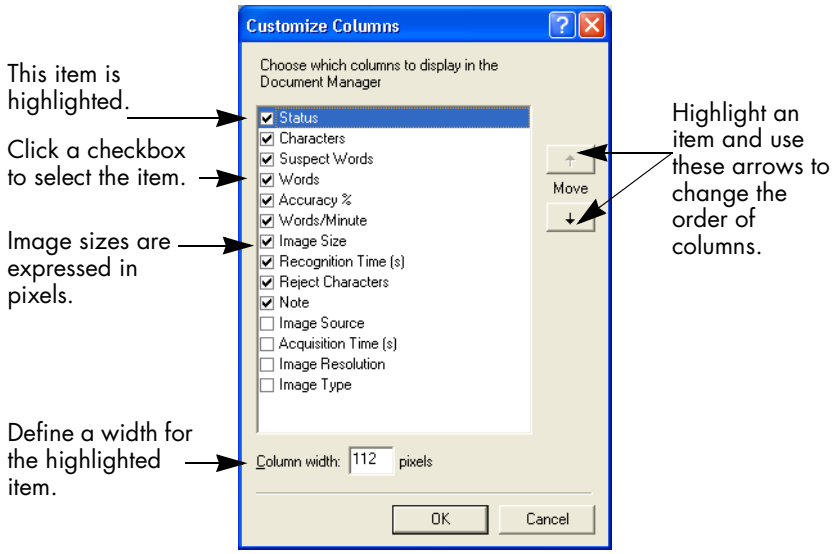
Delete a page: Select the row of the page you want to delete and press the Delete key.

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. You can also export selected pages.

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

Customizing Document Manager columns

You can specify which columns of information you want to see in the Document Manager. Click Customize Columns... in the View menu for the following dialog box:



Define which columns should appear, their widths, and column order. The topic *Customizing Document Manager columns* in online Help clarifies what is presented in each column. You can change column widths easily in the Document Manager; 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 or from the thumbnails 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 tool in the Standard toolbar prints images or text, depending whether the Image Panel 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 training data when closing document' in the Proofing panel of the Options dialog box.

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 Pro 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 (deskewed and pre-processed) 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, 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 Pro. 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.

Recognition results should be saved from OPD files before installing any OmniPage Pro 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. When you open an OPD created by OmniPage Pro 10, only images are loaded. When you open an OPD created by OmniPage Pro 11, images and recognized pages are loaded, but no zones are retained.

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.

The title bar shows the file name of the most recent whole-document save.

Settings

The Options dialog box is the central location for OmniPage Pro settings. Access it from the Standard toolbar or the Tools menu. Context-sensitive help provides information on each setting. In overview, the settings panels are:

OCR
Use this to specify recognition languages, a user or professional dictionary, a reject character and font matching. Click the checkbox before a language to select or deselect it. Multiple selection is possible; select only languages appearing in the document to be recognized. The top items are the recently selected languages. Key in the first letters of a language to jump to it.

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. See “Input from scanner” on page 49. This panel is not available if you requested display of your scanner’s native TWAIN interface when you set up your scanner. See “Setting up your scanner with OmniPage Pro” on page 14.

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. See “How to set up Direct OCR” on page 45.

Process
Use this to define where new images should be placed in the document, to request prompting for more pages when scanning, to specify two-page scanning for handling books, and other settings. 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. See “Proofreading OCR results” on page 65.

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. See “Describing the layout of the document” on page 51.

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. See “Text and image editing” on page 72.

If you have access to a Document Management System (DMS) from your computer and your edition of OmniPage Pro 12 includes ODMA support, an ODMA panel may also appear. See “ODMA support” on page 93.

Some settings have an effect only on future recognition. Examples are the recognition languages, a training file or 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 re-recognize them. Other settings are implemented immediately in all existing pages. Examples are Text Editor settings like word wrap or measurement units.

Chapter 3

Processing documents

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

- Quick Start Guide
- Processing overview
- Automatic processing
- Manual processing
- Combined processing
- Processing with the OCR Wizard
- Processing from other applications (Direct OCR, PaperPort)
- Processing with Schedule OCR

The detailed topics are:

- Defining the source of page images
- Describing the layout of the document
- Zones and backgrounds
  - Automatic zoning
  - Manual zoning
  - Zone types and properties
  - Working with zones
- 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” on page 48. 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 Pro’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.

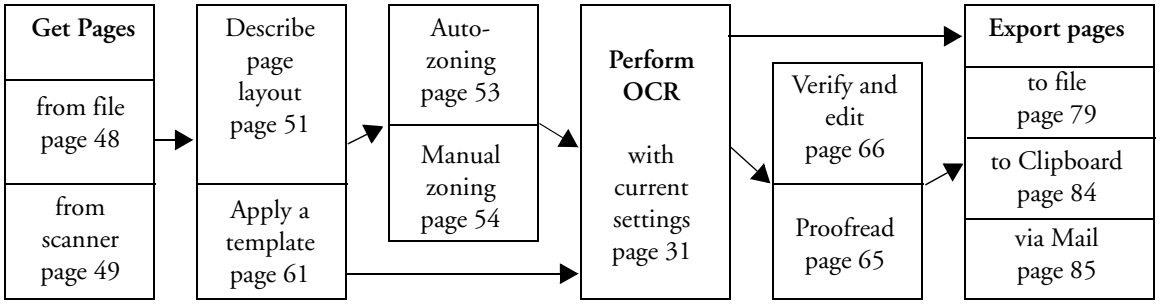
You will process the document automatically and save the recognition results to a file. You will proof the document but will not edit it inside the Text Editor.

	What you do:	What happens:
1.	Set up your scanner using the Scanner Wizard, if this is not already done.	Configures OmniPage Pro to work with your scanner.
2.	Select Start ▸ Programs ▸ ScanSoft OmniPage Pro 12.0 ▸ OmniPage Pro 12.0	Opens OmniPage Pro on your computer.
3.	Place the document correctly in your scanner.	
4.	From the Get Page drop-down list, 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.
5.	From the Layout Description drop-down list, check Automatic is selected. For a wide range of documents, this is the best choice.	Configures the program how to place zones on the page and decide their properties automatically.
6.	From the Export Results drop-down list, check that Save as File is selected.	This means you will be able to name your export file after you have proofed the document.
7.	Click the Start button.	OmniPage Pro will start to scan in your document. A thumbnail appears with a progress indicator. The OCR Proofreader appears.
8.	Use the OCR Proofreader 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. It removes markings from words you proof.
9.	Click in the Text Editor. Select Text Editor views one after another, to see how the page appears in each view.	Each Text Editor view defines a formatting level. This guides you which level to choose at saving time.
10.	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.
11.	Choose a file name, file type, path and a formatting level 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.
12.	Inspect the document in your word processing program.	You have successfully used OmniPage Pro 12 to recognize your document and open it in your target application!

If you succeeded 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” on page 49. This provides a model of optimum brightness. See also the online Help topics *Setting up your scanner* and *Scanner troubleshooting*.

Processing overview

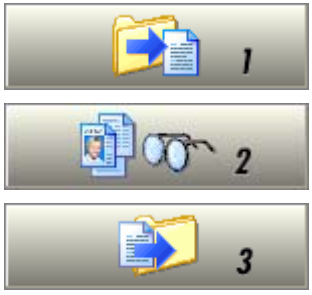
The following flow diagram summarizes the processing steps:



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.

Automatic

The fastest and easiest way to process documents is to let OmniPage Pro do it automatically for you. Select settings in the Options dialog box and in the OmniPage Toolbox drop-down lists 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.



Manual

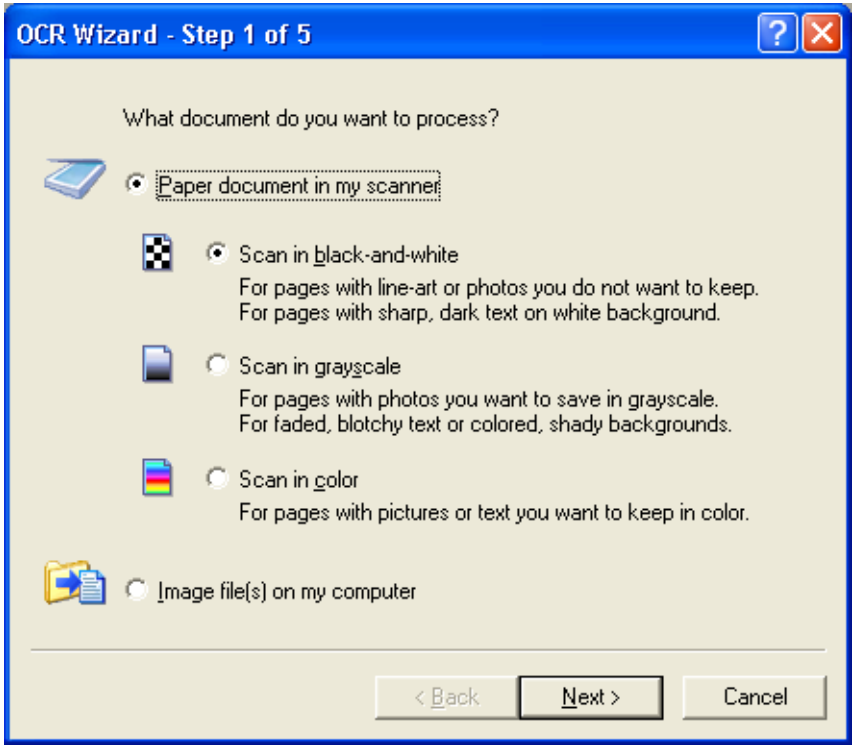
Manual processing gives you more precise control over the way your pages are handled. You can process the document page-by-page with different settings for each page. The program also stops between each step: acquiring images, performing recognition, exporting. This lets you, for instance, draw zones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

Combined

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 re-recognize just those problem pages. Alternatively, you can acquire images with manual processing, draw zones on some or all of them, and then send all pages to automatic processing.

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 Pro.



In other applications

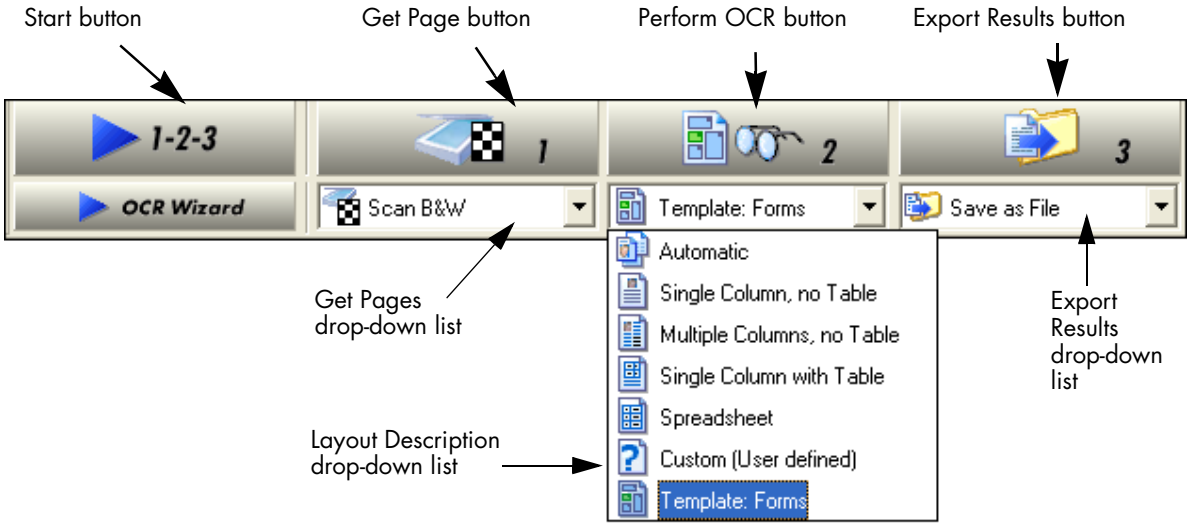
You can use the Direct OCR feature to call on the recognition services of OmniPage Pro while working in your usual word-processor or similar application. OmniPage Pro also automatically links itself to ScanSoft's PaperPort and Pagis document management programs.

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 New Job Wizard in Schedule OCR allows you to specify settings and a starting time.

Automatic processing

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



1. Select the desired Get Page setting in the drop-down list. You define the document source, which can be from image files or from a scanner. See “Defining the source of page images” on page 48.

2. Select a setting from the Layout Description 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. See “Describing the layout of the document” on page 51.

3. Select a setting from the Export Results drop-down list. You can save the document as an OmniPage Document file. You can save pages (current, selected, all) to file, copy them to Clipboard or send them as mail attachments. See “Saving and exporting” on page 77.

4. Choose in the Standard toolbar or 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" on page 31.

5. Click the Start button or choose Start auto-processing 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.

Stopping and restarting automatic processing

Stop: When automatic processing is in progress, the Start button becomes Stop. Click it to interrupt automatic processing. You may do this if you find that some settings need to be changed.

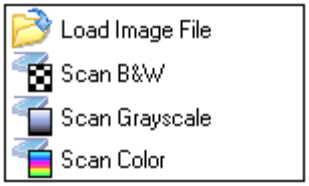
Restart: When automatic processing is stopped, the Start button is restored. Click it to restart processing. The Automatic Processing dialog box lets you specify what you want to do:

- **Finish** processing unrecognized and unproofed pages and then export the results.
- **Export** an already saved 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** to discard all recognition results and re-recognize all pages in the document with different settings. You can specify auto-zoning or a template file. You may want to do this 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. This option lets you perform re-recognition without having to scan or load or rezone all the images again.

Manual processing

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, change the page background and draw zones manually on each page. You start each step in the process by clicking the three numbered buttons on the OmniPage Toolbox.

1. Click in the Standard toolbar or Options in the Tools menu to check or make settings in the Options dialog box. See “Settings” on page 31.

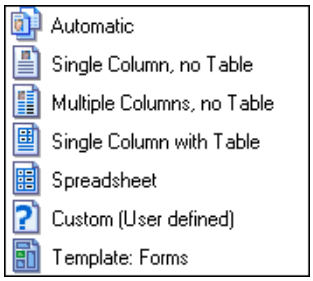


2. Select the desired value for the Get Page button from the drop-down list. You define the document source, which can be from image files or from a scanner. When scanning, select a scanning mode and use the Scanner and Process panels of the Options dialog box to select settings. See “Defining the source of page images” on page 48.

3. Click the Get Page button. This either brings up the Load Image File dialog box allowing you to name images files, or initiates scanning. Thumbnail images of each page can appear in the Image Panel, along with the current page image. Use status bar buttons to show or hide either of these. Acquired pages are summarized in the Document Manager.

4. All page images enter the program with a process background. Provided you draw no zones on these pages, they will be auto-zoned when recognition is requested.

5. You can manually draw and modify zones on one or more images and assign zone properties. Status bar buttons let you move to other pages. As soon as you draw a zone on a page, it takes on an ignore background. You can specify auto-zoning on parts of a page by drawing process zones. See “Zones and backgrounds” on page 53.



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. See “Describing the layout of the document” on page 51. You can also select a template to have its zones placed on the current page. See “Using zone templates” on page 61.

7. Click the Perform OCR button to have the current page recognized. To have selected pages recognized, make a multiple selection with the thumbnails or in the Document Manager (See “Managing documents” on page 26) and then click the Perform OCR button. Recognized pages appear in the Text Editor.

8. 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” on page 65.

9. Continue loading pages, performing OCR, editing, proofing and verifying as desired. You can change the reading order of page elements in the Text Editor. See “Text and image editing” on page 72.

10. Select a value for the Export Results button. You can save the document as an OmniPage Document file. You can save pages (current, selected or all) to file, copy them to Clipboard or send them as mail attachments. Click the Export Results button. See “Saving and exporting” on page 77.

Combined processing

Automatic processing provides speed and efficiency. Manual processing demands more attention, but gives greater control over results. It is possible to tap into both benefits while processing a single document.

Start automatically and finish 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. After adjusting settings and/or modifying zones, use manual processing to re-recognize 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 and delete or modify the existing zones in the Image Panel. You can also load a template to let its zones replace existing ones. Draw new zones as desired. See "Zones and backgrounds" on page 53.

4. Change other settings as required for the current page. See "Settings" on page 31.

5. Click the Perform OCR button to re-recognize the current page. Confirm that the previous recognition results should be overwritten. Alternatively, you can use on-the-fly processing to handle zoning changes without re-recognizing the whole page. See "On-the-fly editing" on page 74.

6. To re-recognize more than one page, select the required pages in the thumbnails or Document Manager before clicking the Perform OCR button.

7. When all pages have been re-recognized with acceptable results, save the document again.

Start manually and finish automatically:

1. Prepare settings and acquire images for the document by clicking the Get Page button.

2. Examine the pages for suitable brightness, orientation and content. Rescan or rotate unsuitable images. Reorder pages as desired.

3. Manually zone pages where you want to process only part of the page or if you want to give precise zoning instructions. Use ignore backgrounds or zones to exclude areas from processing. Use process backgrounds or zones to specify areas to be auto-zoned.

4. Click the Start button, then choose *Finish Processing Existing Pages* in the Automatic Processing dialog box.

5. After proofing (if requested) you can save or export the document.

Processing with the OCR Wizard

The OCR Wizard can be used to start processing a new document. If you select it with a document open, it will be closed. The Wizard takes you through five settings panels, guiding you to make settings for your document and then launching automatic processing. Context-sensitive help is available for all Wizard panels. Click the OCR Wizard button in the OmniPage Toolbox to see the first wizard screen:

1. The first panel lets you define your document source: scanner or image file. See "Defining the source of page images" on page 48. Answer the question 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. See "Describing the layout of the document" on page 51.

3. The third panel lets you define recognition languages. Languages with dictionary support have an open book icon. Recent choices are at the top of the list.

4. The fourth 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 "IntelliTrain" on page 70.

5. 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.

6. If you requested proofing and the text contains suspect words, the OCR Proofreader dialog box will appear. When proofing is finished or closed, the Copy to Clipboard or Save As dialog box let you specify file export settings, including a page range and a formatting level.

7. The document remains in OmniPage Pro. You can edit recognition results and save them again to other formats. You can change zones manually or change other settings and then use manual processing to re-recognize single pages from the document. You can add pages with automatic or manual processing.

The Wizard panels present settings as they were last set in the program. Also, OmniPage Pro 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 click *Start* in the OmniPage Toolbox.

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 from other applications

You can use the Direct OCR feature to call on the recognition services of OmniPage Pro 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 Pro. Start OmniPage Pro, open the Options dialog box at the Direct OCR panel and select *Enable Direct OCR*.

2. Select process options for proofing and zoning. These function for future Direct OCR work until you change them again; they are not applied when OmniPage Pro is used on its own.

3. The Unregistered panel displays running or previously registered applications. Select the desired one(s) and click Add. You can browse for an unlisted application.

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 target application's 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 Pro. Settings changed for Direct OCR are also changed in OmniPage Pro.

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 continue with recognition.

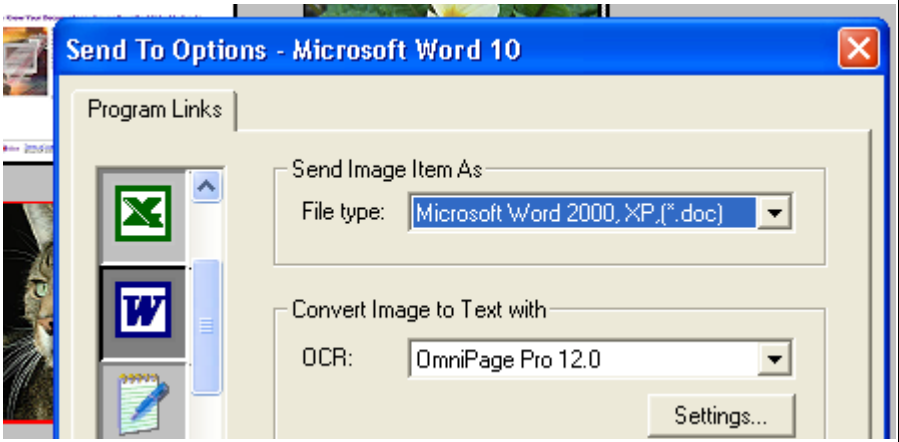
6. If proofing was specified, this follows recognition. Then the recognized text is placed at the cursor position in your application, with the formatting level specified by Acquire Text Settings... .

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

See the Direct OCR topics in online Help for more information. These include a topic *Direct OCR Questions and Answers*. The Readme file and the ScanSoft web site may present more recent information relating to specific target applications.

How to use OmniPage Pro with PaperPort

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 with OmniPage Pro, its OCR services become available and amplify the power of PaperPort. You can choose an OCR program by right clicking on a text application's PaperPort link, selecting Preferences and then selecting OmniPage Pro 12 as the OCR package. OCR settings can be specified, as with Direct OCR.



Here OmniPage Pro 12 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 a PaperPort toolbar. While the text is being recognized, only a progress monitor is displayed. OmniPage Pro'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 with Schedule OCR

You can schedule OCR jobs to be performed automatically at any time within the following eight days. The job 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 Pro 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 Pro 12.0 ▸ Schedule OCR.
2. The Schedule OCR dialog box appears. Click New... to get the New Job Wizard. It takes you through six panels, similar to the OCR Wizard.
3. In the first panel you define image source: scanner with ADF or file.
4. The next two panels are similar to those in the OCR Wizard, but you can also specify a user or professional dictionary and a training file. Whether IntelliTrain runs or not depends on the setting in OmniPage Pro at job time.
5. The following panels let you specify an export file name, type, location, a file separation choice and a formatting level.
6. The last panel lets you define the job start time and (where applicable) a stop time, and retain or delete input files after processing. Click Finish to close the Wizard.

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

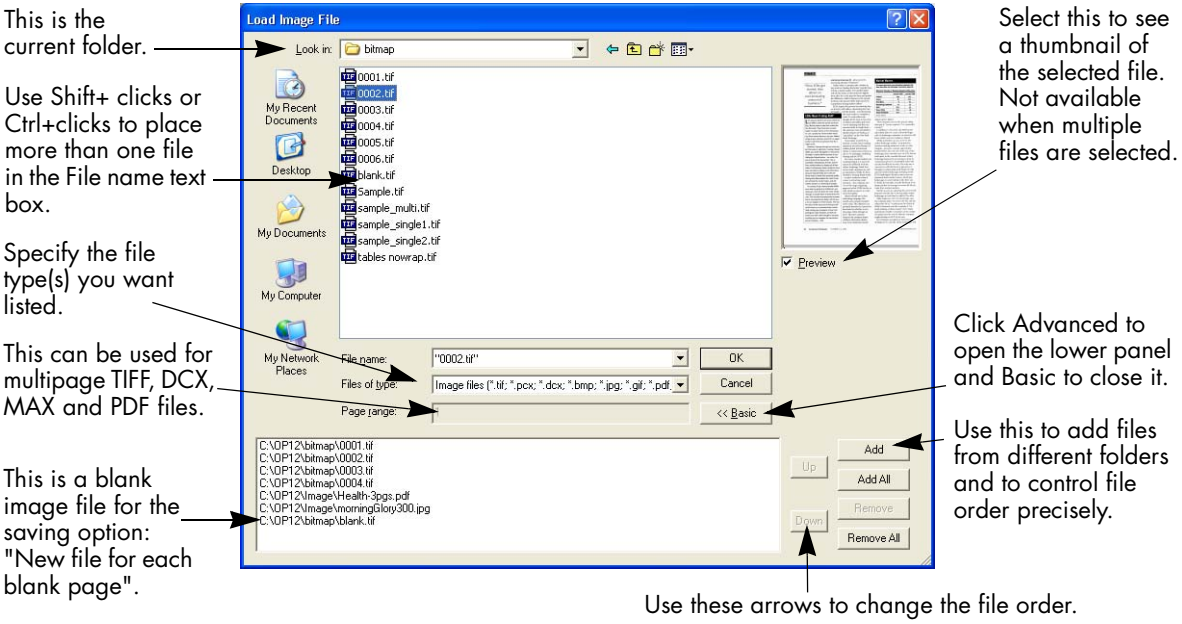
For more information, please see *Scheduling OCR* in the online Help. If your edition of OmniPage Pro supports file input from folders, including watched folders, see “Advanced features in Schedule OCR” on page 93.

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 Pro or may be saved with the scanner's own software to an image file, which OmniPage Pro 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 Pro can open a wide range of image file types. See "File types for opening and saving images" on page 94. Select *Load Image File* in the Get Pages drop-down list. Files are specified in the Load Image File dialog box. This appears when you start automatic processing. In manual processing, click the Get Page 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.



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 Pro. See "Setting up your scanner with OmniPage Pro" on page 14. 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. This will function 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 or in your scanner's interface. 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.

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 recognition results" on page 79.

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 Pro will merge the pages for you.

Scanning 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.

To scan books two pages at a time, select *Look for facing pages* in the Process panel of the Options dialog box. The program will split the incoming images into two pages and deskew them independently.

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 that does not contain a process zone or background. When you do manual processing, auto-zoning sometimes runs. See the online Help topic *When does auto-zoning run*? 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 selected at export time) or without frames (if Flowing Page is selected). 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 background value, zones and properties applied to all acquired pages from now on. The template zones are also applied to the current page, replacing any existing zones. They will also be applied to pre-existing pages without zones when they are (re-)recognized. See “Using zone templates” on page 61.

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

Zones and backgrounds

Zones define areas on the page to be processed or ignored. Zones are rectangular or irregular, with vertical and horizontal sides. Page images in a document have a background value: process or ignore (the latter is more typical). Background values can be changed with the tools shown. Zones can be drawn on page backgrounds with the tools shown:

Backgrounds	Process	Ignore			
Zones	Process	Ignore	Text	Table	Graphic

Process areas (in process zones or backgrounds) are auto-zoned when they are sent to recognition.

Ignore areas (in ignore zones or backgrounds) are dropped from processing. No text is recognized and no image is transferred.

Automatic zoning

Automatic zoning allows the program to detect blocks of text, headings, pictures and other elements on a page and draw zones to enclose them. It assigns zone types and properties to those zones. Auto-zoning runs on whole pages when you do automatic processing, unless you have a template loaded. It runs when you use the OCR Wizard. You can also specify auto-zoning when doing manual processing, as follows:

Auto-zone a whole page
Acquire a page. It appears with a process background. Draw no zones on it and check in the Layout Description drop-down list that a zone template is not loaded. Click the Perform OCR button. You can select several zone-less pages to have them auto-zoned and recognized together.

Auto-zone a part of a page
Acquire a page. It appears with a process background. Draw a zone. The background changes to ignore. Draw text, table or graphic zones to enclose areas you want manually zoned. Draw process zones to enclose areas you want auto-zoned. After recognition the process zones will be replaced with one or more text, table or graphic zones.

Auto-zone a page background

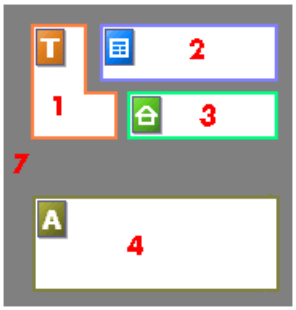
Acquire a page. It appears with a process background. Draw a zone. The background changes to ignore. Draw text, table or graphic zones to enclose areas you want manually zoned. Click the Process background tool (shown) to set a process background. Draw ignore zones over parts of the page you do not need. After recognition the page will return with an ignore background and new zones round all elements found on the background.

Manual zoning

First we present two examples on zones and backgrounds. Then we detail the zone types. Lastly we explain how to draw and work with zones. In these examples the numbers refer to the table on the following page.

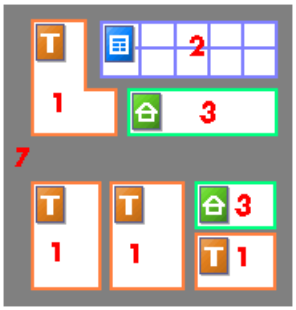
Drawing zones on an ignore background:

Before recognition:



After recognition:

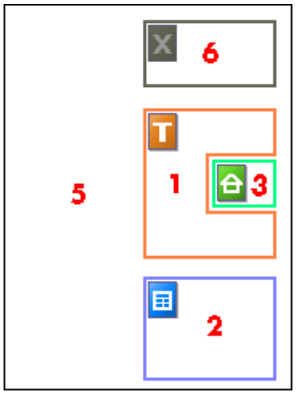
Background remains as ignore.



Zone 4 returns as a set of zones, in this case to handle three columns of text and a photo.

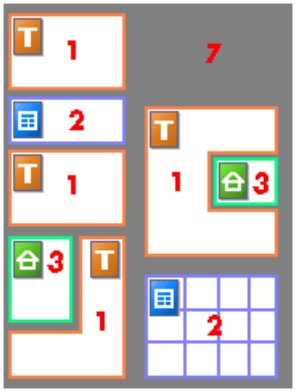
Drawing zones on a process background:

Before recognition:



After recognition:

Background is changed to ignore.



Zone 6 is absorbed into the background. All zones on the left side of the page were automatically created.

No.	Type	What happens:
1	Text zone	OCR runs and generates text.
2	Table zone	OCR runs, text is placed in a table grid.
3	Graphic zone	Image is embedded in recognized page.
4	Process zone	Auto-zoning creates one or more zones, decides their types and processes their contents.
5	Process background	
6	Ignore zone	Nothing
7	Ignore background	

Automatically drawn zones and template zones have solid borders:

Manually drawn or modified zones have dotted borders:

Zones do not have a reading order. Reordering of recognized elements can be done in the Text Editor. See "Text and image editing" on page 72. On-the-fly zoning is described in chapter 4. See "On-the-fly editing" on page 74.

Zone types and 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 with Shift+clicks to change their properties in one move.

The Image toolbar provides five zone drawing tools, one for each type. A zone's type is shown by an icon in its top left corner, and by the icon and zone border color. Here are the tools and the colors:

Process zone (olive)

Use this to draw a process zone, to define a page area where auto-zoning will run. After recognition, this zone will be replaced by one or more zones with automatically determined zone types. You normally draw

process zones on an ignore background. Draw a process zone to enclose columns of text to have them handled automatically. They will be decolumnized in the Text Editor's NF view and RFP view, but kept in columns in True Page view.

Ignore zone (gray)

Use this to draw an ignore zone, to define a page area you do not want transferred to 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 an ignore zone in a template. You normally draw ignore zones on a process background.

Text zone (brown)

Use this to draw a text zone. Draw it over a single block of text. Zone contents will be treated as flowing text, without columns being found. If you want columns of text to be handled automatically, enclose them in a process zone.

Table zone (blue)

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. For many output file types, you can choose whether to export tables in grids or in columns separated by tabs.

Graphic zone (green)

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. Embedded images can be exported with the document to target applications supporting graphics.

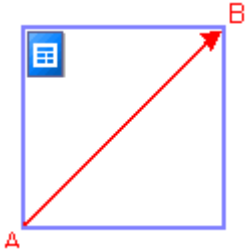
Text and table zones have a zone content setting. Alphanumeric contents validates all characters needed for your language choice. Recognition results from a numeric zone will contain only numbers and number-related punctuation. No letters will be placed. Use the zone's shortcut menu to change this setting.

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.

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. Some tools on this toolbar are grouped. Only the last selected tool from the group is visible. To select a visible tool, click it. To select a hidden tool, hold down the mouse button on the triangle at the bottom right of the visible tool until the additional tools appear, then click the tool you want.

Draw a single zone



Select the zone drawing tool of the desired type, then click and drag the cursor. In these examples, this is shown by the arrow going from A to B. Dragging from top left to bottom right is also possible.

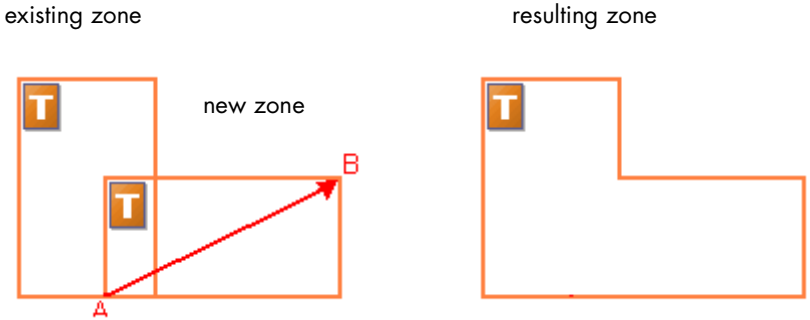
Only rectangular zones can be drawn; zones (except table zones) can be made irregular after they are drawn.

To **resize** a zone, select it by clicking in it, move the cursor to a side or corner, catch a handle and move it to the desired location.

To **move** a zone, select it with the zone selection tool and move it as desired. You cannot move a zone to overlap another zone.

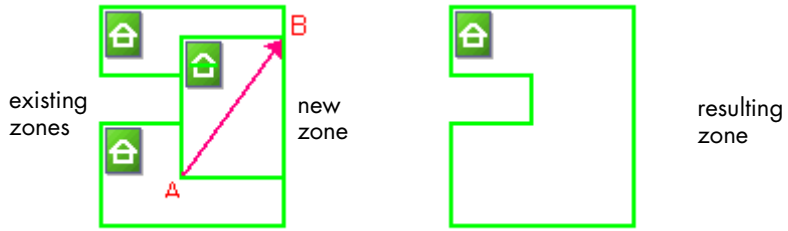
Make an irregular zone by addition

Draw a partially overlapping zone of the same type:



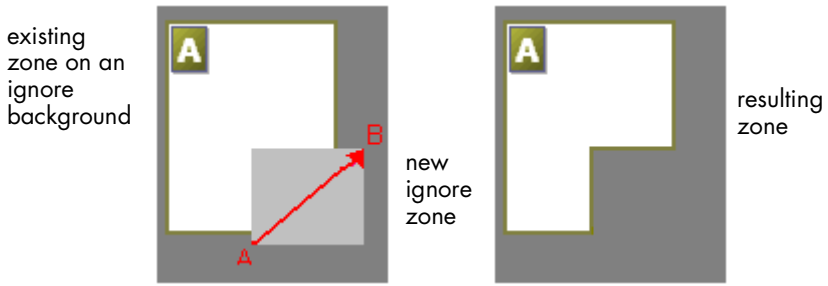
Join two zones of the same type

Draw an overlapping zone of the same type.



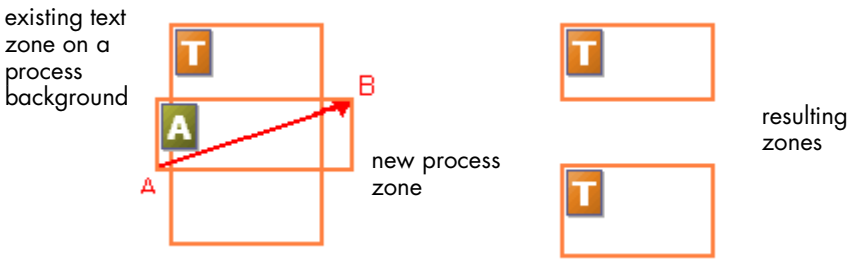
Make an irregular zone by subtraction

Draw an overlapping zone of the same type as the background (in this example, on an ignore background).

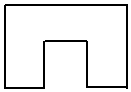


Split a zone

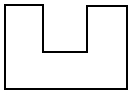
Draw a splitting zone of the same type as the background (in this example, on a process background).



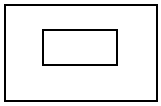
The following zone shapes are prohibited:



Indented along the bottom



Indented along the top



Hole in the middle

To expand a zone more quickly than using its resizing handles, draw a zone of the same type to completely enclose it. The smaller zone is replaced by the larger one. To replace a set of zones of whatever type with a single zone, draw a larger zone of the desired type to completely enclose them. All the smaller zones are replaced by the larger one.

When you draw a new zone that partly overlaps an existing zone of a different type, it does not really overlap it; the new zone replaces the overlapped part of the existing zone.

Diagrams in the online Help topic *Drawing zones manually* clarify these two topics.

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 left corner of the zone. To change a rectangular zone to or from a table zone, use its shortcut menu. You can also draw table type zones, but they must remain rectangular.

You draw or move table dividers to determine where gridlines will appear when the table is placed in the Text Editor. You can draw or resize a table zone (provided it stays rectangular) to discard unneeded columns or rows from the outer edges of a table.

The five grouped table handling tools on the Imaging toolbar can be used if the current page contains a table type zone. If the tool you need is not visible, click the triangle on the bottom right of the visible tool to display all the tools, then click the desired one.

Use the table tools and their cursors as follows:

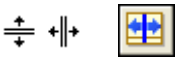
Insert row dividers

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

Insert column dividers

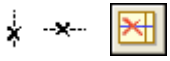
Click the tool then click at the location in a table zone where you want to place a column divider.

Move dividers



Click the tool and move the cursor to the row or column divider to be moved. It displays a double-headed arrow. Drag the divider as desired. You cannot drag it beyond its neighbor. Avoid placing dividers so they cut through text.

Remove dividers



Click the tool then click on a single row or column 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.

Place/Remove all dividers

Click this tool and click its cursor icon inside a table zone without dividers. Dividers will be auto-detected and placed. Click it in a table with dividers to make them all disappear.

Press the Ctrl key as you click if you want to place, move or delete a divider in the current cell only.

You can specify line formatting for table borders and grids from a shortcut menu. You will have greater choice for editing borders and shading in the Text Editor after recognition.

Using zone templates

A template contains a page background value and a set of zones and their properties, stored in a file. A zone template file can be loaded to have template zones used during recognition. Load a template file in the Layout Description drop-down list or from the Tools menu.

When you load a template, its background and zones are placed:

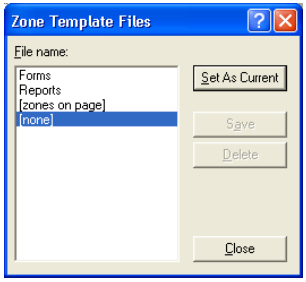
- on the current page, replacing any zones already there
- on all further acquired pages
- on pre-existing pages sent to (re-)recognition without any zones.

With manual processing the template zones in the first two cases can be viewed and modified before recognition.

With automatic processing the template zones can be viewed and modified only after recognition.

This behavior continues until the template is unloaded.

Templates accept ignore and process zones and backgrounds. They can therefore be useful to define which parts of the pages to process with auto-zoning, and which parts to ignore. Process zones or process background areas from a template may be replaced during recognition by a set of smaller zones; specific zone types will be assigned to these zones.



How to save a zone template
Select a background value and prepare zones on a page. Check their locations and properties. Click Zone Template... in the Tools menu. In the dialog box, select [zones on page] and click Save, then assign a name and click OK.

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 in the Layout Description 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 Layout Description drop-down list, or open the Zone Template Files dialog box, select the desired template and click the Set As Current button. Zones from the new template are applied to the current page, replacing any existing zones. They are also applied as explained above.

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.

Templates are available in Schedule OCR and Direct OCR, but not in the OCR Wizard.

Chapter 4

Proofing and editing

Recognition results are placed in the Text Editor. These can be recognized texts, tables and embedded graphics. This WYSIWYG (What You See Is What You Get) editor offers the following features, detailed in this chapter:

- The editor display and views
- Proofreading OCR results
- Verifying text
- User dictionaries
- Training
- Text and image editing
- On-the-fly editing
- Reading text aloud

The editor display and views

The Text Editor displays recognized texts and can mark words that were suspected during recognition with wavy underlines:

- **Green – Non-dictionary words:** These were recognized confidently, but are not found in any active dictionary: standard, user or professional.
- **Blue – Words with suspect characters:** These contain unrecognized characters or are dictionary-approved words containing characters recognized with lower confidence.
- **Red – Suspect words:** These are likely to be non-dictionary words with one or more suspect characters, but may also be suspect for other reasons.

Choose to have non-dictionary words marked or not in the Proofing panel of the Options dialog box. All 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.

OmniPage Pro 12 can display pages with three levels of formatting. You can switch freely between them with the three buttons at the bottom left of the Text Editor or from the View menu. Graphics and tables can appear in all views. 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

True Page® 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 boxes and frames. Reading order can be displayed by arrows. See from page 72.

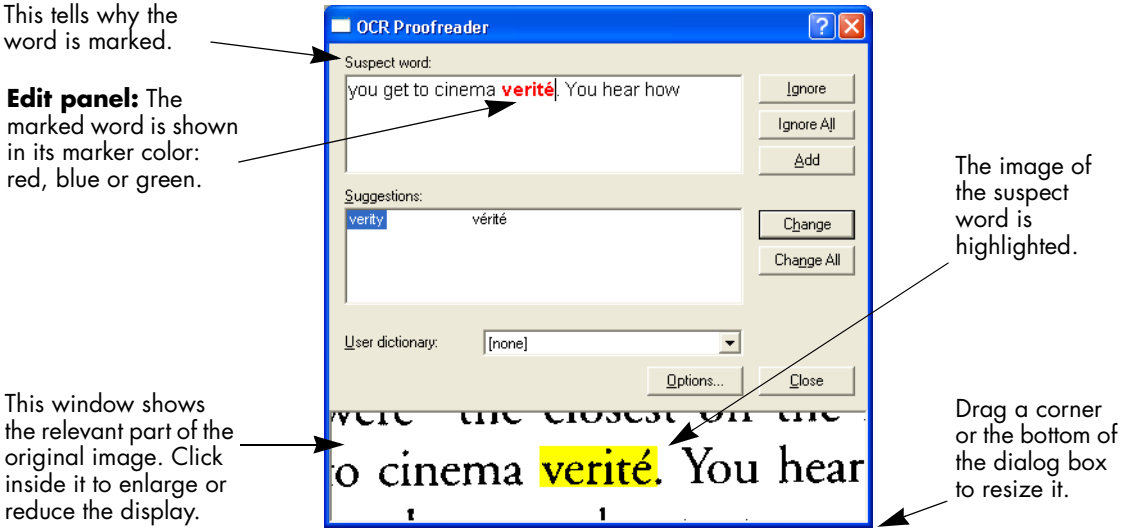
The formatting level for export is chosen separately at export time.

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. Work as follows:

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

2. Proofing starts from the current page, but skips text already proofed. If a suspected error is detected, the OCR Proofreader dialog box colors the suspect word in its context, and provides a picture of how it originally looked in the image.



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, modify the word in the Edit panel 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 changed word 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.

A page is marked with the proofed icon on its thumbnail and in the Document Manager if proofing ran to the end of the page.

If markers were hidden in the Text Editor when proofing is started or *Find Next Suspect* is chosen, the markers become shown and remain shown after proofing.

If *Mark non-dictionary words* is turned off in the Proofing panel of the Options dialog box, proofing will stop only on words marked red or blue, and not on non-dictionary words. This is useful when checking pages with many non-dictionary words, such as product catalogues containing codes and bibliographies containing many proper names.

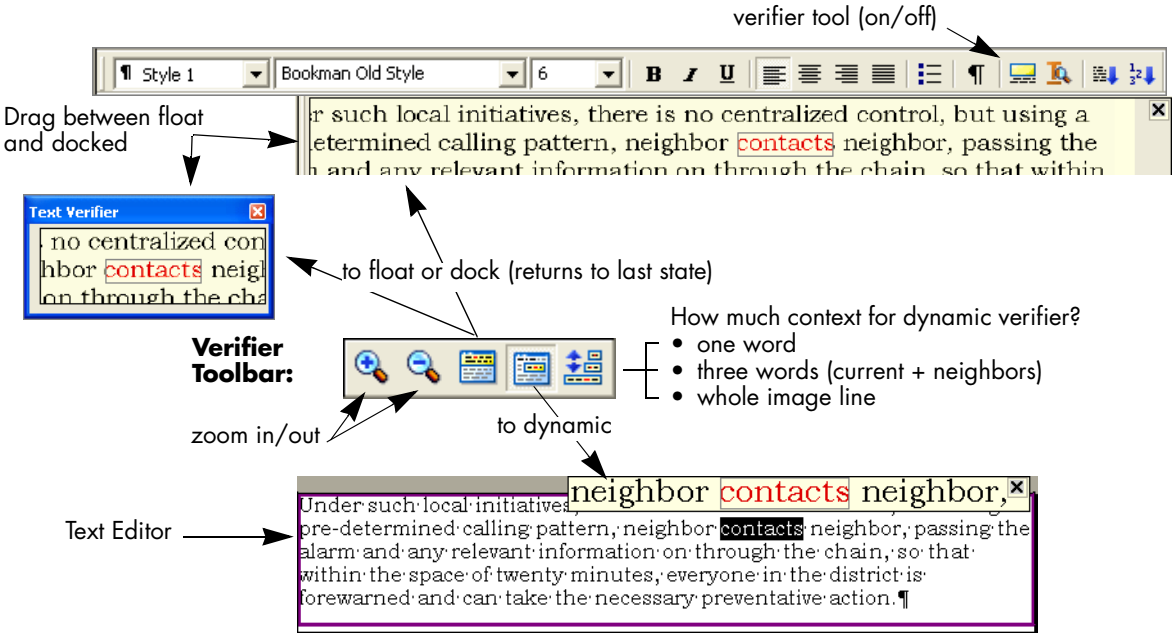
Use *Recheck Current Page* in the Tools menu to run a new spelling check on a page that has already been proofed. Do this to check words typed or pasted in the Text Editor after proofing was done. This works even if *Mark non-dictionary words* is turned off in the Proofing panel.

Verifying text

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:

To do this:	Use this:
Turn verifier on	F9 or verifier tool
Turn verifier off	Esc or F9 or verifier tool
Turn verifier on/off temporarily	F8: press and hold down
Show verifier until next keystroke	Double-click on word
Zoom display in	Alt + Num + or click in verifier
Zoom display out	Alt + Num – or click in verifier
Make verifier dynamic or docked/floating	Alt + Num /
Dynamic context (scroll through 3 values)	Alt + Num *

The verifier tool is in the Formatting toolbar. The verifier can also be controlled from the Tools menu. Hover the cursor over a verifier display to obtain the verifier toolbar. Use it as follows:



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.

You can use OmniPage Pro's Text-to-Speech facility to have the recognized text read 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 the section "Reading text aloud" on page 75.

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 or deleting 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 by clicking Edit in the User Dictionary Files dialog box. The Delete button lets you delete the selected user dictionary.

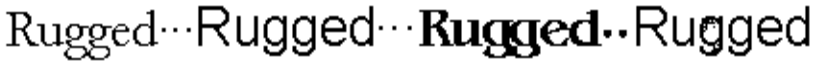
While editing a user dictionary, you can import a word list from a plain text file to add words to the dictionary quickly. Each word must be on a separate line with no punctuation at the start or end of the word.

In addition to user dictionaries, specialized dictionaries are available for certain professions (currently medical and legal) for some languages. See the list and make selections in the OCR panel of the Options dialog box.

The program identifies the language of recognized texts and displays it in the status bar. This language marking is exported with the document. Use Set Language... in the Tools menu to change the language marking for selected text. This does not change the recognition language(s).

Training

Training is the process of changing the OCR solutions assigned to character shapes in the image. It is useful for uniformly degraded documents or when an unusual typeface is used throughout a document. Training 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:



The first two examples do not need training, because both shapes are normal for the letter “g” and the program can handle them. The third example could benefit from training 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 training, because the first “g” is poorly printed, and this shape is unlikely to appear again in the document.

You can use training to improve recognition of special symbols such as @, ® and © or to recognize supported accented letters more reliably. The purpose of training is not to teach the program to read characters from non-supported languages or alphabets.

OmniPage Pro 12 offers two types of training: manual training and automatic training (IntelliTrain). Data coming from both types of training are combined and available for saving to a training file.

When you leave a page on which training data was generated, you will be asked how to apply it to other existing pages in the document.

Manual training

To do manual training, place the insertion point in front of the character you want to train, or select a group of characters (up to one word) and choose Train Character... from the Tools menu or the shortcut menu. You will see an enlarged view of the character(s) to be trained, along with the current OCR solution. Change this to the desired solution and click OK. The program takes this training and examines the rest of the page. If it

finds candidate words to change, the Check Training dialog box lists these. Incorrect words should be re-trained before the list is approved.

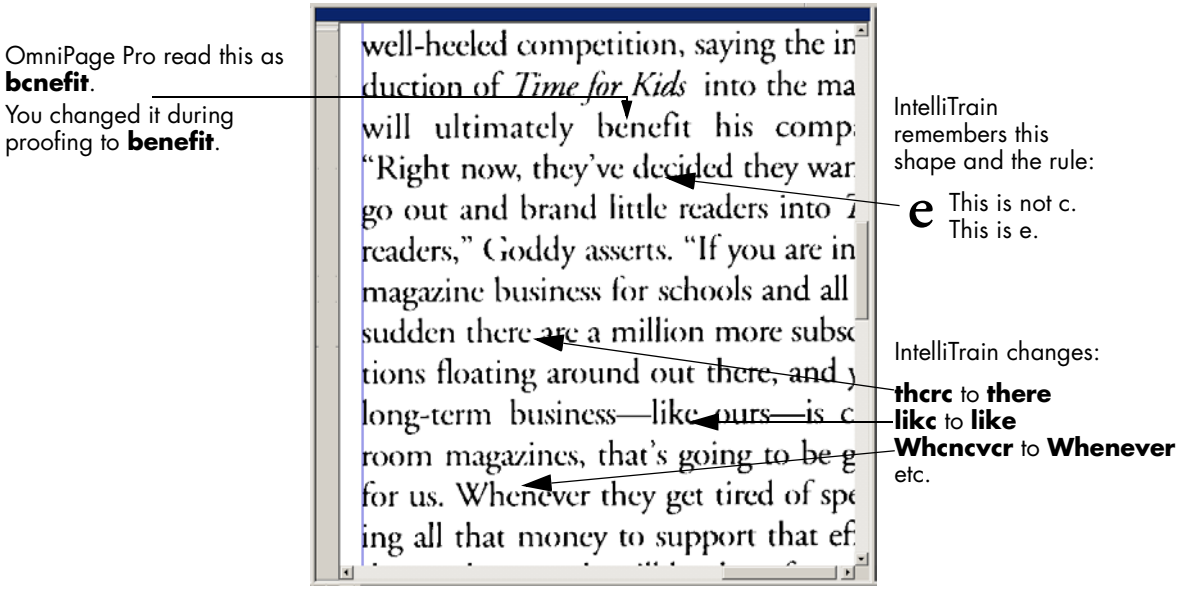
For guidance on using the Train Character and Check Training dialog boxes, please consult their context-sensitive help or the online help topic *Manual training* and its related topics.

IntelliTrain

IntelliTrain is an 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.

You can turn IntelliTrain on or off in the OCR panel of the Options dialog box.

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.

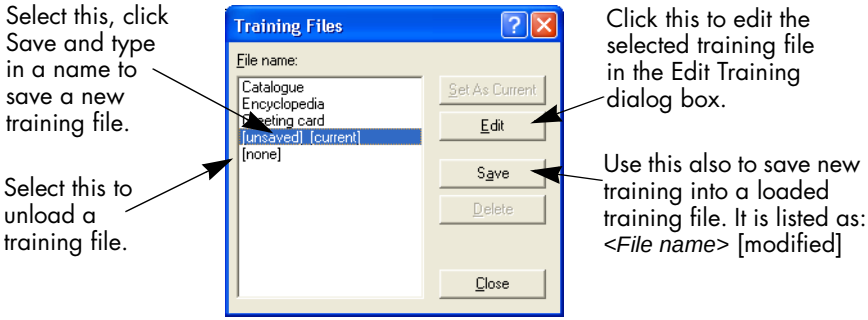


IntelliTrain remembers the training data it collects, and adds it to any manual training you have done. This training can be saved to a training file for future use with similar documents.

Training files

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. If you do not save the document as an OPD, unsaved training is discarded when the document is closed.

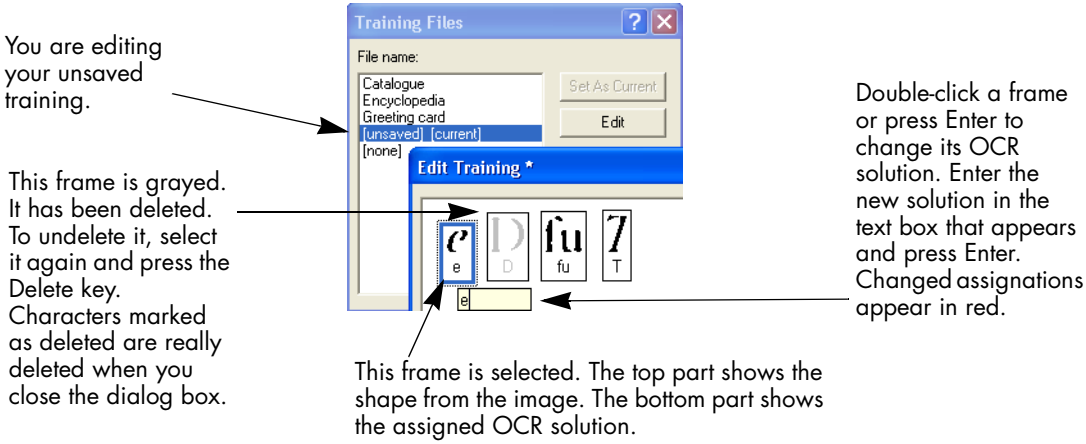
Saving training to file, loading, editing and unloading training files are all done in the Training Files dialog box. Open this from the Proofing panel of the Options dialog box or the Tools menu.



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 Edit Training dialog box. Save it in the Training Files 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 Edit Training 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.



Text and image editing

OmniPage Pro has 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, use the Formatting toolbar to specify one font type and size to be applied to the whole document. This is not used for export, nor transferred to other views; their previous settings are restored.

Open the Font Matching dialog box from the OCR panel of the Options dialog box before OCR, 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 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 re-embedded in the Text Editor. Do not change the graphic's size, resolution or type, because this will prevent the re-embedding.

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 some file types, you can choose whether to have the tables exported in grids or as tab separated or space separated columns.

Hyperlinks
Web page and e-mail addresses can be detected and placed as links in recognized text. Choose Hyperlink... in the Format menu to edit an existing link or create a new one. A new link can be to a web page or a file. Use a shortcut menu to delete a link.

Editing in True Page
Page elements are contained in text boxes, table boxes and picture boxes. These usually correspond to text, table and graphic zones in the image. Click inside an element to see the box border; they have the same coloring as the corresponding zones. The online Help topic *True Page* provides details on the operations summarized here.

Frames have gray borders and enclose one or more boxes. They are placed when a visible border is detected in an image. Format frame and table borders and shading with a shortcut menu or by choosing Table... in the Format menu. Text box shading can be specified from its shortcut menu. To call up a shortcut menu, right-click inside an element away from a marked word.

Multicolumn areas have pink borders and enclose one or more boxes. They are auto-detected and show which text will be treated as flowing columns when exported with the Flowing Page formatting level. Use shortcut menus to ungroup multicolumn areas and frames, allowing their elements to be modified. You can also group elements into frames or multicolumn areas.

Reading order can be displayed and changed. Click the Show reading order tool in the Formatting toolbar to have the order shown by arrows. Click again to remove the arrows. Click the Change reading order tool for a set of reordering buttons in place of the Formatting toolbar. Context-sensitive help explains their use, as does *Reading order* in online Help. A changed order is applied in NF and RFP views. It modifies the way the cursor moves through a page when it is exported as True Page.

On-the-fly editing

This allows you to modify a recognized page through re-zoning, without having to re-process the whole page. When on-the-fly editing is enabled, zone changes (deleting, drawing, resizing, changing type) immediately make changes in the recognized page. Conversely, when you modify elements in the Text Editor's True Page view, this changes the zones on that page. On-the-fly zoning can also be used with unrecognized pages.

Two linked tools on the Image toolbar control on-the-fly zoning. One of these tools is always active whenever no recognition is in progress.

Click this to activate on-the-fly editing. The red signal shows there are no stored zoning changes.

Click this to turn on-the-fly editing off. Your zoning changes are stored; the on-the-fly tool displays a green signal to show there are stored changes. To activate these changes, do one of the following:

Click the on-the-fly tool with a green signal. The zoning changes will cause changes in the Text Editor.

Click the Perform OCR button to have the whole page (re)recognized, including your zone changes.

For details on how changes are handled in on-the-fly zoning and their effects in the Text Editor views, see *On-the-fly processing* in online Help.

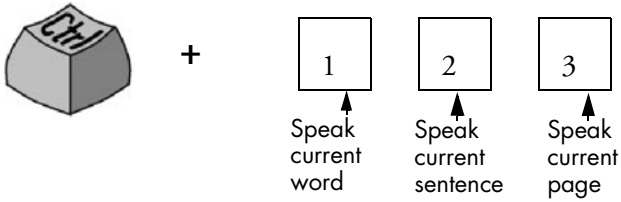
Reading text aloud

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.

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.

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.

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 SAPI-compliant speech systems for other languages on your computer, they will be detected and available. Their voices will be offered 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.

Chapter 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 – a single page, selected pages or the whole document – to a target application by saving to file, copying to Clipboard or sending to a mailing application. Saving as an OmniPage Document is always possible.

This chapter presents the following topics:

- Saving original images
- Saving recognition results
  - Saving a document as you work
  - Selecting a formatting level
  - Selecting advanced saving options
  - Saving to PDF
- Copying pages to Clipboard
- Sending pages by mail

A document remains in OmniPage Pro after export. This allows you to save, copy or send its pages repeatedly, for example with different formatting levels, using different file types, names or locations. You can also add or re-recognize 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), an exporting dialog box appears.

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 re-recognize those pages before exporting.

Saving original images

You can save original images to disk in a wide variety of file types. See "File types for opening and saving images" on page 94.

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 zone image only, the current page image, selected page images or all images in the document. In the last two cases you can have all images in a single multi-page image file, providing you set TIFF, MAX or DCX as file type. Otherwise each image is placed in a separate file. OmniPage Pro 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.

To see the image size and original resolution of an image, hover the cursor over its thumbnail in the Image Panel.

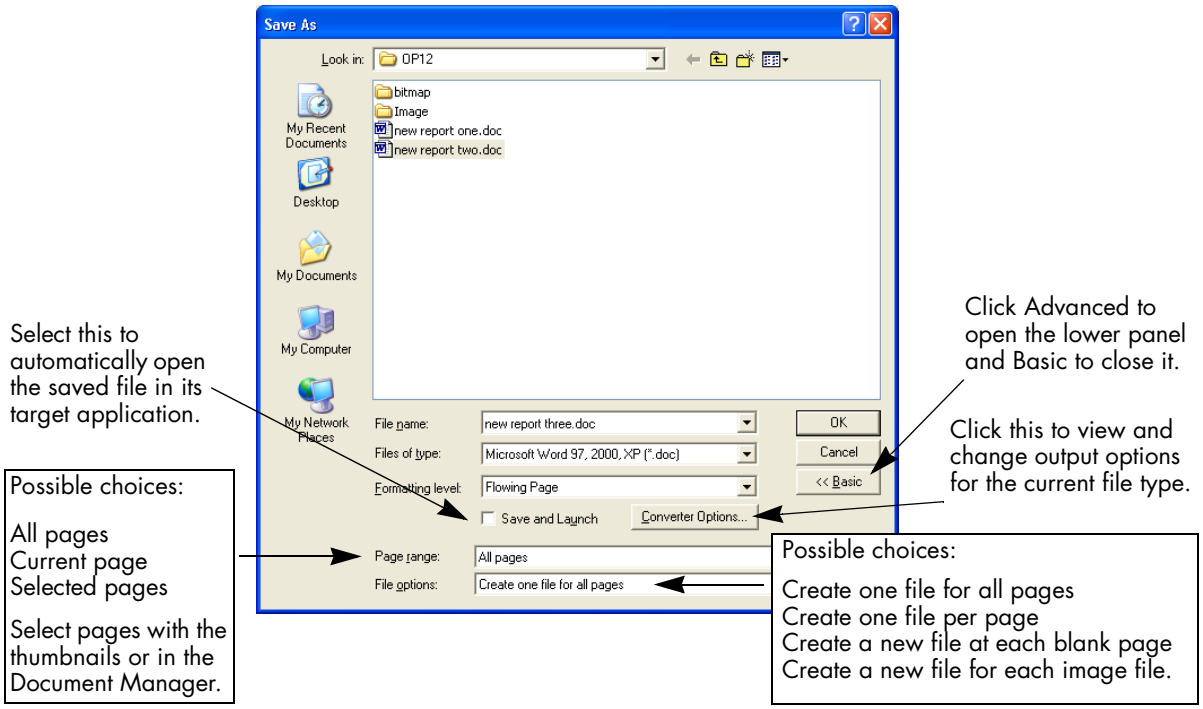
You can save your document to five variants of PDF. Two of these save the original images, the others save recognition results. See the following sections.

Saving recognition results

You can save recognized pages to disk in a wide variety of file types. See "File types for saving recognition results" on page 95.

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

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



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. Then select a formatting level for the document. See "Selecting a formatting level" on page 81.

4. Type in a file name. Click the Advanced button if you want to specify a page range, a file separation option or other saving options. Select these as desired. See "Selecting advanced saving options" on page 82.

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 or in the advanced saving options for your selected file type converter.

Graphics, table grids and other properties are saved in the document only if the selected file type supports them, and if these are specified for retention in the advanced settings (Converter options) for the current file type.

If more than one export file is created, OmniPage Pro will append numerical suffixes 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, you can place blank image files in the document. See "Input from image files" on page 48.

If you select *Create a new file for each image file*, no file name is required. Each output file will take its name from the input file that generated it, with just the extension changed.

Saving a document as you work

Click the Save tool 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 file type of its last full save, as displayed in the title bar. These display only if the whole document has been saved.

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 any other file type, only changes to the recognition results are saved.

If you want to work with your document again in OmniPage Pro 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 "OmniPage Documents" on page 29.

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.

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.

Selecting a formatting level

The formatting level for export is defined at export time, in the Save As dialog box, the Copy to Clipboard dialog box or the Send as Mail dialog box. Three of the levels correspond to the format views of the same name in the Text Editor. However, the level to be applied for saving is independent of the formatting view displayed in the Text Editor. When exporting to file or mail, first specify a file type. This determines which formatting levels are available. A table in chapter 6 summarizes this. See "File types for saving recognition results" on page 95.

The formatting levels are:

No Formatting (NF)
This exports plain decolumnized left-aligned text in a single font and font size. When exporting to Text or Unicode file types, graphics and tables are not supported. You can export plain text to nearly all file types and target applications; in these cases graphics, tables and bullets can be retained.

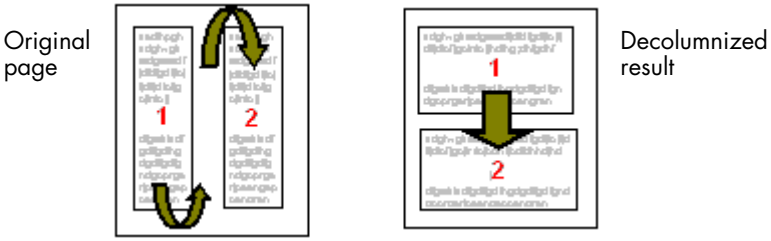
Retain Fonts and Paragraphs (RFP)
This exports decolumnized text with font and paragraph styling, along with graphics and tables. This is available for nearly all file types.

Flowing Page (FP)
This keeps the original layout of the pages, including columns. This is done wherever possible with column and indent settings, not with text boxes or frames. Text will then flow from one column to the other, which does not happen when text boxes are used.

True Page (TP)
This keeps the original layout of the pages, including columns. This is done with text, picture and table boxes and frames. This is offered only for target applications capable of handling these. True Page formatting is the only choice for XML export and for all PDF export, except to the file type 'PDF Edited'.

Spreadsheet
This exports recognition results in tabular form, suitable for use in spreadsheet applications.

Decolumnization for NF and RFP export is performed from left-to-right and top-to-bottom:



Before export, check in NF or RFP view that the decolumnized order of elements is correct. If not, switch to True Page view and click the Show reading order tool to have the order shown by arrows. Use the Change reading order tool to specify a different order. Multicolumn areas show which columns are linked. If this linking is unsuitable, ungroup the area and change the order of the elements it enclosed.

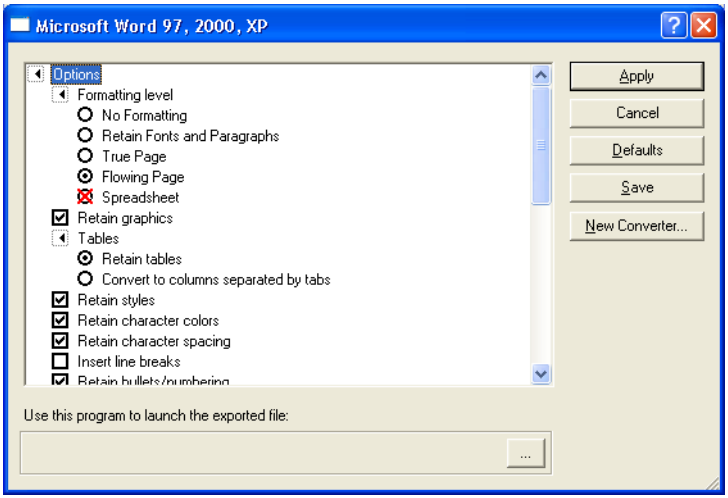
Selecting advanced saving options

Click the Converter Options button in the advanced part of the Save As dialog box to have precise control over the export. This brings up a dialog box with the name of the current file type. It presents a series of options tailored to this file type. First, confirm or change the formatting level, because this influences which other options are presented. Select options as desired. Online Help details how to do this.

Click Apply to have the changed settings applied to the current save only.

Click Defaults to have all settings returned to the default values for the current file type.

Click Save to have the changed settings applied to the current save and also stored as the settings to be applied in future whenever this file type is selected again for saving.



The program currently associated with the chosen file type for the Save and Launch feature is displayed at the bottom of the dialog box. Click the three dots button to specify a different program.

To make your own customized converter, prepare your settings, click New Converter, provide a name, then click OK. Alternatively, name the converter first, change settings next and then click Save. Custom converters are useful for repeated tasks, such as publishing a weekly magazine. Then all recognized pages can be exported with their formatting tailored to their intended use. You can also create a set of customized converters for a given file type defining saving options for each output formatting level, for example: RTF No Formatting, RTF Retain Fonts and Paragraphs, RTF Flowing Page and RTF True Page.

You can change converter options without saving anything to file. Call the Export Converters dialog box from the Tools menu. Select the desired converter and click the Options button. In this case, the Apply button is not available.

Saving to PDF

You have five choices when saving to Portable Document Format (PDF) files.

PDF (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.

PDF Edited:
Use this if you have made significant editing changes in the recognition results. You have three formatting level choices, including True Page. The PDF file can be viewed, searched and edited.

PDF with image on text:
The PDF file is viewable only and cannot be modified in a PDF editor. The original images are exported, but there is a linked text file behind each image, so the text can be searched. A found word is highlighted in the image.

PDF with image substitutes:
As for PDF (Normal), but words containing reject and suspect characters have image overlays, so these uncertain words display as they were in the original document. The PDF file can be viewed, searched and edited.

PDF, image only:
The original images are exported. The PDF file is viewable only and cannot be modified in a PDF editor and text cannot be searched.

Copying pages to Clipboard

You can copy the recognition results from the current page, selected pages or all document pages 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 6.0/95 information. Otherwise, only

plain or Unicode text will be pasted. Graphics are retained if the application supports insertion of images.

▼ **To copy pages to the Clipboard:**

- With automatic processing, select Copy to Clipboard as the setting in the Export Results drop-down list on the OmniPage Toolbox or in the OCR Wizard. The Copy to Clipboard dialog box appears as soon as the last available page is recognized or proofed.
- With manual processing, select the Copy to Clipboard setting in the Export Results drop-down list and then click its button. The Copy to Clipboard dialog box appears immediately.
- Specify a page range and formatting level to be used, then click OK to start the copying.

You can perform a copy and paste operation for the current page by drag-and-drop. Press and hold Ctrl+Tab as you click the current page in the Image Panel and drag the cursor to a target application with an open document. The page contents will be pasted at the cursor position. OCR runs if necessary.

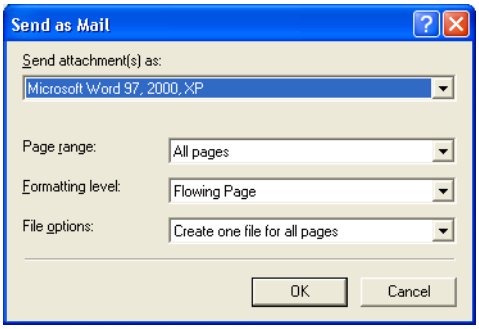
Sending pages by mail

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 pages by e-mail:**

- With automatic processing, select Send as Mail as the setting in the Export Results drop-down list on the OmniPage Toolbox. The Send as 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 setting in the Export Results drop-down list and then click its button. The Send as 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 as Mail dialog box.



1. This dialog box lets you specify a file type, a page range, a formatting level and attachment options: one attachment for all pages, one attachment per page, new attachment at each blank page or one attachment for each input file. 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 Pro, 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.

The program can detect e-mail addresses as it recognizes pages and transmits these to the Text Editor. If you click an address, your mailing application appears with a new empty message containing only the e-mail address.

Chapter 6

Technical information

This chapter provides troubleshooting and other technical information about using OmniPage Pro 12. Please also read the online Readme file and other help topics, or visit the ScanSoft web pages. Its scanner section contains detailed and regularly updated information about scanner setup and support. The Readme file contains last-minute information relating to OmniPage Pro. 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 Pro
  - Increasing memory resources
  - Increasing disk space
  - Text does not get recognized properly
  - Problems with fax recognition
  - System or performance problems during OCR
- ODMA support
- Advanced features in Schedule OCR
- Supported file types
  - File types for opening and saving images
  - File types for saving recognition results
- Uninstalling the software

Troubleshooting

Although OmniPage Pro 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 Pro:

- Make sure that your system meets all the listed requirements. See "System requirements" on page 12.
- 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 Pro 12. 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 Pro.
- Make sure you have the correct drivers for your scanner, printer, and video card. Visit ScanSoft's web page through the Help menu and consult its scanner section for more information.
- Run ScanDisk for Windows 98 or Me, or Check Disk for Windows NT, 2000 and XP 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 Pro, as described in the last section, "Uninstalling the software" on page 96.

Testing OmniPage Pro

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

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

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

1. Restart your computer in safe mode by pressing F8 immediately after you see the 'Starting Windows' message.
2. Launch OmniPage Pro and try performing OCR on an image. Use a known image file, for instance one of the supplied sample image files.
   - If OmniPage Pro does not launch or run properly in safe mode, then there may be a problem with the installation. Uninstall and reinstall OmniPage Pro (see the end of this chapter), and then run it in Windows safe mode.
   - If OmniPage Pro runs in safe mode, then a device driver on your system may be interfering with OmniPage Pro operation. Troubleshoot the problem by restarting Windows in Step-by-Step Confirmation mode. See Windows online Help for more information.

▼ **To test OmniPage Pro 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 Pro and try performing OCR on an image. Use a known image file such as one of the supplied sample files.

You can also run OmniPage Pro 12 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.

Increasing memory resources

OmniPage Pro may run poorly under low-memory conditions. This may be indicated by various error messages or if OmniPage Pro 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 Pro 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" on page 12.

Increasing disk space

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 or Check Disk.
- 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.
- Clear the cache for your web browser and limit its size.

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 the section on creating and modifying zones, "Working with zones" on page 57.
- Make sure text zones are identified correctly. Reidentify zone types and contents, if necessary, and perform OCR on the document again. See "Zone types and properties" on page 55.
- 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 a page thumbnail 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.
- Do some manual training, or edit existing training to remove unsuccessful training.
- If you use True Page as the Text Editor view or for export, recognized text is put into text boxes or frames. Some text may be hidden if a text box is too small. To view the text, place the cursor in the text box and use the arrow keys on your keyboard to scroll to the top, bottom, left, or right of the box.
- Check the glass, mirrors, and lenses on your scanner for dust, smudges or scratches. Clean if necessary.

OmniPage Pro only recognizes machine printed-text characters such as type-written 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 Pro. See "Input from image files" on page 48.

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 and low disk space problems. See "Testing OmniPage Pro" on page 89.
- Minimize all applications or click Alt+Tab to check for Windows error messages.
- Check image quality. 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" on page 57.
- Restart Windows 98, Me, 2000 or XP in safe mode, or Windows NT in VGA mode and test OmniPage Pro by performing OCR on the included sample image files.

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

ODMA support

If your local network includes a Document Management System (DMS) that supports ODMA clients, OmniPage Pro may be able to work with it. Then an ODMA panel will appear in the Options dialog box allowing you to specify permissible file types and other settings. An ODMA interface will replace the Load Image File and Open OmniPage Document (OPD) dialog boxes. This lets you load image files and OPDs one at a time from the network file system or your local computer. The Save As dialog box will provide a Save to DMS button for saving recognized documents into that system. For detail, see online Help.

Advanced features in Schedule OCR

Schedule OCR allows you to specify input files for a job. Some editions of OmniPage Pro let you specify that all files of a given type in one or more folders be processed. These editions also offer watched folder jobs. The first New Job Wizard screen has two buttons: Files and Folders. It also displays a checkmark *Watch folders for incoming files*. Select this and specify one or more folders. Then all files of the specified type(s) entering the folder(s) will be processed on arrival. In the fifth wizard panel you can specify both a start and stop time for the job watching process. You can choose not to specify a stop time when you set up the job. In that case, use the Pause and Modify buttons to enter a stop time later.

You can pause and resume jobs. The View button lets you see a file-by-file log of all completed processing inside a selected job. When a job is running, a job log window is available, displaying file-by-file progress and reporting any processing problems.

The fourth New Job Wizard panel lets you specify a file type and choose file separation options. If you choose *A new output file for each input file,* you specify only a folder; output files retain their input names with modified extensions. If you choose an option requiring multiple output files, you supply one file name and the program appends numerical suffixes to generate unique file names. If you specify input from a set of folders, you can specify a different output folder for each input folder.

Supported file types

The program supports a wide range of file types for images and text.

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	No	Open and Save	All
JPEG	jpg	No	Open and Save	Grayscale, color
MAX	max	Yes	Open and Save	All
PCX	pcx	No	Open and Save	All
PDF	pdf	Yes	Open *(see note)*	All
PNG	png	No	Open and Save	All
TIFF Compressed G3	tif	Yes	Open and Save	B/W
TIFF Compressed G4	tif	Yes	Open and Save	B/W
TIFF Compressed LZW	tif	Yes	Open and Save	All
TIFF FX	xif	Yes	Open	All
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. That means these are not good candidates for future OCR processing.

Hover the cursor over a page thumbnail for a popup window showing the size and resolution of the original image.

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.

Saving to PDF format is supported, with five options. Two of these, *Image only* and *Image on text*, export original images. This is done in the Save As dialog box for recognized pages. See “Saving to PDF” on page 84.

File types for saving recognition results

This table shows which formatting levels are available for each file type.

File type	Extension	No Formatting	RFP	Flowing Page	True Page	Spread sheet	Graphics	Tables
eBook (1)	opf	●	●				●●	●
Excel 97, 2000	xls	●	●			●	●●	●
Excel 3.0 to 7.0	xls	●	●			●		●
FrameMaker 5.5.3	mif	●	●				●●	●●
Freelance Graphics	txt	●	●					❍●
Harvard Graphics	txt	●	●					❍●
HTML 4.0 (2)	htm	●	●		●		●●	●
HTML 3.2 (2)	htm	●	●			●	●●	●
Microsoft PowerPoint 97	rtf	●	●				●●	●●
Microsoft Publisher 98	rtf	●	●				●●	●●
Microsoft Word 6.0, 97, 2000, XP	doc	●	●	●	●		●●	●●
PageMaker 6.5.2	doc	●	●				●●	●●
Quattro Pro for Windows 4.0, 8	xls	●	●			●		●
PDF (Normal)	pdf				●		●●	●
PDF Edited	pdf	●	●		●		●●	●●
PDF with image on text	pdf				(●)		●	●
PDF with image substitutes	pdf				●		●	●
PDF, image only	pdf				(●)		●	●
RTF Word 2000, 97, 95/6.0 (3)	rtf	●	●	●	●		●●	●●
Ventura Publisher	doc	●	●				●	●●
WordPad	rtf	●	●				●●	●●
WordPerfect 8, 9, 10	wpd	●	●	●	●		●●	●●
WordPerfect 6.0, 6.1	wpd	●	●		●		●●	●●
WordPerfect 5.1,5.2	wp5	●	●		●		●●	●●
XML (1)	xml				●		●	(●)
Text and Text with line breaks (4)	txt	●	●					❍●
Text – Comma Separated (4)	csv	●	●					●
Text – Formatted (4)	txt				●			(●)
OmniPage Document (5)	opd	Saved as displayed					●	●

Graphics

● File type supports graphics

●● File type supports graphics, with export choice to retain or drop graphics.

Tables

● File type supports tables in grids, no table handling choices at export time

●● File type supports tables, choose to use grids or tab separated columns

❍● File type does not supports table grids, choose to convert to tab or space separated columns

1 These new formats are available only in some editions of the program.

2 When saving to HTML, all graphics are saved as separate JPEG image files.

3 Recognition results are sent to Clipboard in RTF 95/6.0 and will be pasted in this format if possible, and as Unicode or ASCII text if not.

4 All text formats are available as Text or Unicode. The latter can handle the widest range of accented characters.

5 OmniPage Documents can be reopened by OmniPage Pro. It can open OPD files created by OmniPage Pro 10 or 11. These files enter the program as unnamed documents. To keep an OPD in the old format and also save it as a new OPD, choose a different name to avoid overwriting the old file.

Uninstalling the software

Sometimes uninstalling and then reinstalling OmniPage Pro will solve a problem. OmniPage Pro's Uninstall program will not remove any of the following user-created files:

Zone templates (*.zon)

Training files (*.otd)

User dictionaries (*.ud)

OmniPage Documents (*.opd)

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

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

- ◆ Close OmniPage Pro.
- ◆ Click Start in the Windows taskbar and choose the Control Panel and then Add/Remove Programs.
- ◆ Select OmniPage Pro and click Change.
- ◆ Click Next in the dialog box that appears.
- ◆ Select Remove All or Repair All, then Next.
- ◆ Follow instructions until the process is finished.

INDEX

A

Accuracy
- improvement, 49, 69, 91
- influence of brightness, 50
- influence of training, 69
- scanning mode influence, 49

Acquire Text menu items, 45
Acquired pages, 26
Acquiring images, 21, 40
Adding
- pages to a document, 39
- to zones, 58
- training to training files, 71
- words to a user dictionary, 66

ADF, 31, 48, 50
Advanced saving options, 82
Advice on problems, 88
Alphanumeric zone, 55
Attachments to mail messages, 85
Auto-detect layout, 51
Automatic Document Feeder (ADF), 31, 48, 50
Automatic processing, 25, 38
Automatic training, 70
Auto-zoning, 24, 32, 38, 51, 56

B

Backgrounds for zoning, 24, 53
Basic processing steps, 21
Black-and-white
- images, 78
- scanning, 49

Bold text, 72
Book scanning, 31
Boxes, 24, 73
Boxes for recognized text, 91
Brightness, 31, 50, 91

C

Changing
- part of a page, 75
- reading order, 74
- zone types, 56

Character attributes, 72
Characters
- suspect, 64
- unrecognized, 64

Checking OCR results, 66
Clipboard, 39, 84
Closing documents, 29
Color
- images, 78
- markers, 66
- scanning, 49

Columns
- in Document Manager, 28
- in tables, 60

Combined processing, 25, 41
Comparing recognized words with originals, 66
Contents of OmniPage Documents, 80
Context-Sensitive Help, 9, 23, 31
Contrast, 31, 50, 91
Control over processing, 40
Conversion of images, 94
Copying pages to Clipboard, 43, 84
Creating training data, 71
Custom Layout, 32, 52
Customizing
- Document Manager columns, 28
- export converters, 82
- toolbars, 23

D

Deferred processing, 29
Deleting
- pages, 26, 28
- training files, 71
- user dictionaries, 68
- zone templates, 61

Describing document layout, 38, 51
Desktop, 22
Dictionaries, 43, 66
Direct OCR, 44
- Options panel, 31

Disk space, 12, 90
DMS support, 93
Docking toolbars, 23
Document Manager, 22, 26, 27
- customizing columns in, 28

Documents
- closing, 29
- copying to Clipboard, 43, 84
- double-sided, 51
- exporting, 21, 38, 41, 77
- finishing, 39
- in OmniPage Pro, 21
- layout description, 51
- managing, 26
- place for new pages, 31
- saving, 30, 77
- saving as you work, 80
- unfinished, 29
- with varied layout, 51

Dot-matrix texts, 92
Double-sided documents, 51
Drawing zones in Direct OCR, 45
Drop-down list
- Export Results, 41
- Get Pages, 40
- Layout Description, 41

Dropping graphics from export, 79
Duplex scanners, 51
Dynamic verifier, 66

E

Editing
- character attributes, 72
- graphics, 73
- in True Page, 73
- on-the-fly, 75
- paragraph attributes, 72
- PDF output, 84
- recognized text, 72
- tables, 59, 73
- training files, 71
- user dictionaries, 68

Effect of settings, 32
Examples of training, 69
Export converters, 82
Export Results button, 39, 41, 79
Exporting
- file types and formatting levels, 95
- Flowing Page, 81
- graphics, 79, 96
- repeated, 77, 80
- to Clipboard, 84
- to file, 79, 95
- to mail, 85
- to PDF, 84, 95

to target applications, 21, 40, 78
True Page, 82

F

Fax recognition, 92
Features, new, 17
Files
as export target, 78
as image source, 48
retained on uninstalling, 96
separation options, 79, 86
types, 79
types for export, 81, 95
types supported, 94
Finding
non-dictionary words, 65
suspect words, 65
Finishing a document, 39
Floating toolbars, 23
Flowing Page, 81
Folder input for Schedule OCR, 93
Formatting levels, 47, 64, 95
Formatting levels and file types, 95
Formatting toolbar, 22, 23
Frames, 24, 73, 82, 91

G

Generating table dividers, 60
Get Page button, 38, 40
Get Pages drop-down list, 40
Getting online Help, 9
Graphic zone, 56
Graphics
editing, 73
in export, 79, 95
in HTML files, 96
Grayscale
images, 78
scanning, 49
Grouping elements, 73

H

Header/footer indicators, 64
Hearing texts read aloud, 76
Help
Context-Sensitive, 9, 23, 31
online, 9
Hiding or showing markers, 64
Hyperlinks, 73

I

Ignore backgrounds, 53
Ignore zones, 56
Image files
input, 20, 48
opening, 94
reading order, 48
samples, 89
types, 94
Image Panel, 22, 24
Image toolbar, 22, 23
Images
acquiring, 21, 40
backgrounds, 53
black-and-white, 78
color, 78
conversion, 94
editing, 73
grayscale, 78
quality, 50
resolution, 27, 78, 91, 94
saving, 78, 94
size, 27
substitutes in PDF, 84
Improving accuracy, 49, 70, 91
Incomplete automatic processing, 39
Increasing disk space, 90
Increasing memory resources, 90
Input
from image file, 48
from PDF files, 48, 94
from scanner, 49
Inserting table dividers, 60
Installing
OmniPage Pro, 13
scanners, 14
IntelliTrain, 32, 47, 70, 91
Interface language, 31
Interrupting automatic processing, 39
Irregular zones, 57
Italic text, 72

J

Jobs in Schedule OCR, 47
Joining zones, 58

K

Keyboard guide for hearing texts, 76

L

Languages
for recognition, 31, 43, 91
for user interface, 31
Launch target application, 80
Layout description, 38, 43, 51
Layout retention, 65
Layout, auto-detect, 51
Legal dictionaries, 66
Links to web pages, 73
Load Image File dialog box, 48
Loading
a user dictionary, 68
OPD files, 29
training files, 71
zone templates, 52, 61
Location for new pages, 31

M

Mail, 39, 85
Managing documents, 26
Manual processing, 25, 40
Manual training, 69
Manual zoning, 40, 53
Marked words in Text Editor, 64
Markers, 64, 66
Medical dictionaries, 66
Memory requirements, 12, 90
Menu bar, 23
Minimum system requirements, 12
Modified pages, 26
Modifying zone templates, 61
Moving
between pages, 26
table dividers, 60
MS Outlook, 85
Multicolumn areas, 24, 73
Multi-page image files, 48, 78, 94
Multiple column pages, 52
Multiple page selection, 26

N

New features, 17
New file on blank page, 48
New Job Wizard, 47, 93
New page placing in document, 31
No Formatting view, 64, 81
Non-dictionary words, 64
Non-printing characters, 64
Note column in Document Manager, 28
Numeric zone, 55

O

OCR
automatic processing, 25, 38
checking OCR results, 66
definition, 20
Direct OCR, 31, 44
jobs in Schedule OCR, 47
manual processing, 25, 40
performing OCR, 21
poor performance during, 92
proofreading results, 65
Schedule OCR, 47
settings, 31
settings for Direct OCR, 44
Wizard, 37, 43, 44

ODMA support, 93
OmniPage Desktop, 22
OmniPage Documents
contents of, 80
definition, 29
purpose of OPD files, 30
saving as, 30, 80
OmniPage Pro
documents in, 21
earlier versions, 13
installing, 13
new features of, 17
registering, 17
reinstalling, 96
starting, 14
testing, 89
uninstalling, 96
OmniPage Toolbox, 22, 25, 38
Online
HTML Help, 9
registration, 17
On-the-fly editing and zoning, 75
OPD files
definition, 29
purpose of, 30
saving to, 30
Opening image files, 48, 94
Optical character recognition, 20
Optimizing brightness, 31, 50
Options dialog box, 31
Options for proofing, 32, 65
Options for saving, 82
Order of page elements, 74
Original image saving, 78, 94
Overview
of document, 26
of processing, 36
of processing steps, 21
Overview of processing, 34

P

Pages
acquired, 26
copying to Clipboard, 43, 84
deleting, 26, 28
Get Page button, 38, 40
location in document, 31
modified, 26
moving between, 26
multi-page image files, 48, 78, 94
multiple column, 52
navigation, 22, 76
new file on blank page, 48
pending, 26
proofed, 26
recognized, 26
reordering, 26
re-recognizing all, 39
saved, 26
selecting multiple, 26
sending as mail, 85
single column, 52, 55
single column pages with tables, 52
spreadsheet pages, 52
status, 26
zoned, 26
PaperPort, 46
Paragraph
editing attributes, 73
retaining paragraph styles, 80
styles, 73, 79
PDF file input, 48, 94
PDF file output, 94
Pending pages, 26, 75
Perform OCR button, 38, 41
Performance problems during OCR, 92
Printing
documents, 29
recognized pages, 29
Problems with fax recognition, 92
Process backgrounds, 53
Process options, 31
Process zones, 56
Processing
automatic, 25, 38
basic steps of, 21
combined, 25, 41
documents in future sessions, 29
from other applications, 44
incomplete auto-processing, 39
interrupting auto-processing, 39
manual, 25, 40
restarting auto-processing, 39
step-by-step, 40
steps, overview, 21, 36
stopping automatic processing, 39
switching between manual and automatic processing, 25, 41
with OCR Wizard, 43
Professional dictionaries, 66
Prompt to save training data, 29
Proofed pages, 26
Proofing
in later sessions, 29
options, 32, 65
Proofreader dialog box, 65
Proofreading OCR results, 65
Properties of zones, 55
Purpose of OPD files, 30
Purpose of training, 69

Q

Quality of images, 50
Quick Start Guide, 34

R

Reading
order of image files, 48
text aloud, 75
Reading order, 74
Recognition
accuracy, 50, 69, 91
languages, 31, 43, 91
performing, 40
problems with fax recognition, 92
saving results, 79
speeding up, 92
Recognized pages, 26
Rectangular zones, 57
Registering
applications for Direct OCR, 45
OmniPage Pro, 17
Reinstalling OmniPage Pro, 96
Remote proofing, 29
Removing table dividers, 60
Reordering pages, 26
Repeated exporting, 77, 80
Replacing zone templates, 61
Re-recognizing pages, 41
Resizing zones, 57
Resolution, 27, 78, 91, 94
Restarting automatic processing, 39
Retain Fonts and Paragraphs view, 64, 81
Retaining paragraph styles, 79
Re-training, 69
Rows in tables, 60

S

Safe mode, 89
Sample image files, 34
Sample images files, 89
Saved pages, 26
Saving
as OmniPage Document, 30, 80
documents, 77
documents as you work, 80
options, 82
original images, 78, 94
recognition results, 79
Save and Launch, 80
text, 79
to file, 44, 78
to OPD format, 30, 79
training files, 71
user dictionaries, 68
zone templates, 61
Scanners, 49, 91
drivers, 14
duplex, 51
setting up, 14

Scanning
black-and-white, 49
books, 31
brightness, 31, 50
color, 49
contrast, 31
grayscale, 49
input from, 49
pictures, 49
Wizard, 14
Schedule OCR, 47
input from folders, 93
watched folders, 93
Searching PDF output, 84
Selecting multiple pages, 26
Send Mail dialog box, 85
Sending pages by mail, 85
Setting up a scanner, 14
Setting up Direct OCR, 45
Settings
Acquire Text, 45
effect of settings, 32
for Direct OCR, 45
in OCR Wizard, 44
in Options dialog box, 31
zone types, 59
Shortcut menus, 56
Single-column
pages, 52, 55
pages with tables, 52
Slow recognition, 92
Solutions for poor performance, 88
Splitting zones, 55
Spreadsheet pages, 52
Standard toolbar, 22, 23
Starting a user dictionary, 68
Starting the program, 14
Step-by-step processing, 21, 40
Stopping automatic processing, 39
Storing zoning changes, 75
Subtracting from zones, 55
Suggestions during proofing, 66
Supported file types, 94
Suspect words, 64
Switching between manual and automatic processing, 25, 41
System or performance problems during OCR, 92
System requirements, 12

T

Tables
columns in, 60
editing, 73
editing dividers, 59
generating dividers, 60
in single column pages, 52
inserting dividers, 60
moving dividers, 60
removing dividers, 59
rows in, 59
table handling in Text Editor, 73
zones, 56, 59
Task Manager, 89
Technical information, 87
Template zones, 52, 61, 91
Testing OmniPage Pro, 89
Text Editor, 22, 24, 32, 64
Text Editor views, 24, 64
Text saving, 79
Text zone, 56
Text-to-Speech facility, 13, 76
Thumbnails, 22, 24, 26
TIFF image files, 94
Toolbar docking and floating, 23, 66
Training, 69
automatic, 70
creating training data, 71
editing training files, 71
IntelliTrain, 70
loading training files, 71
manual, 69
prompt to save data, 29
saving training files, 71
training files, 71
unloading training files, 71
unsaved training data, 29
Troubleshooting, 87, 88
True Page, 24
True Page editing, 73
True Page export, 82
True Page view, 65
TWAIN drivers for scanners, 14
Two-page scanning, 31
Types of zones, 55

U

Underlined text, 72
Unfinished documents, 29
Ungrouping elements, 73
Uninstalling the software, 96
Unit of measurement, 32
Unloading a user dictionary, 68
Unloading training files, 71
Unloading zone templates, 61
Unsaved training data, 29
URLs, 73
User dictionaries, 66, 68
adding words, 66
editing, 68
loading, 68
starting, 68
unloading, 68
Using Direct OCR, 45

V

Verifying text, 66
VGA mode, 89
Views
No Formatting, 64
Retain Fonts & Paragraphs, 64
True Page, 65

W

Watched folders, 93
Web page links, 73
Wizard
for processing, 43
for scanner setup, 14
for Schedule OCR, 47, 93
Word wrapping, 32
Working with zones, 57

Z

Zones, 24
adding to, 58
alphanumeric, 55
changing types, 56
deleting templates, 61
drawing in Direct OCR, 45
graphic, 56
ignore, 56
irregular, 57
joining, 58
manual, 53, 91, 92
modifying templates, 61
numeric, 55
on page, 26
process, 56
properties, 55
rectangular, 57
replacing templates, 61
resizing, 57
saving templates, 61
setting types, 60
splitting, 56
subtracting from, 56
table, 56, 59
templates, 52, 61, 91
text, 56
types, 24, 55, 91
unloading templates, 62
working with, 57
Zoning on-the-fly, 75
Zooming displays, 22, 66