OmniPage Pro

Version 6 for Windows

Reference Manual

How to Use the Documentation

Use this online Reference manual to find specific information about any OmniPage feature. It describes all the commands and settings, how to use True Page, how to improve performance, and how to troubleshoot common problems.

This information is also available in OmniPage's online Help system. Additionally, chapter 2 contains a variety of tutorial exercises to introduce you to basic scanning and many features of OmniPage.

Use the toolbar buttons in Adobe® Acrobat® Reader to view the Bookmarks or Thumbnails. Clicking on the Bookmarks or Thumbnails navigates to the paragraphs or pages of the OmniPage Reference manual.

Assumptions and Symbols

We assume that you know how to work in the Microsoft Windows environment. If you have questions about how to use dialog boxes, scroll bars, edit boxes, and so on, please refer to the Windows User's Guide.

This symbol means ***Note***. It introduces a tip or an item of note.

This symbol means ***Warning***. It introduces cautionary text.

CAERE CORPORATION
100 Cooper Court
Los Gatos, California 95030-3321

European Offices:
CAERE GmbH
Innere Wiener Strasse 5
81667 Munich, Germany

OmniPage Pro Version 6 *(PC Windows version)*

Chapter 1

Installation

Please read this section carefully! It includes:

- What's in the Package
- System Requirements
- Saving a User Dictionary Before Installation
- Setting up a Windows Swap File
- Installing the Software
- Starting OmniPage Pro

What's in the Package

Your OmniPage Pro 6.0 package includes:

- Omnipage Pro program
- OmniPage Pro online manual, which can be printed
- OmniPage Reference manual, if separately requested

System Requirements

To install and run OmniPage, you need the following setup:

- Computer with an 80386 or higher processor.
- Microsoft Windows version 3.1 or higher.
- Windows-compatible mouse.
- Total system memory of at least 8MB RAM.
  12MB RAM are recommended for Windows for Workgroups users.
- 8MB or larger permanent Windows swap file.
- Super-VGA color monitor with 512K memory on the adapter card.

To view all 24 bits of color (millions of colors) in 24-bit color images, you need a 24-bit video card.

- A compatible scanner if you plan to scan documents.
  See the list of supported scanners in the Release Notes.

Install your scanner and test it according to the manufacturer's instructions before using it with OmniPage.

Saving a User Dictionary Before Installation

Read this section if you have a user dictionary from an older version of OmniPage. OmniPage 6.0 overwrites the user dictionary (`user.ud`) if you install the program in the `omnipro` directory.

If you are upgrading from OmniPage 5.x, move the 5.x user dictionary to another directory before installation. This preserves your entries. Move the 5.x user dictionary back to the `omnipro` directory after installation to overwrite the newer user dictionary.

A user dictionary from OmniPage or OmniPage Professional 2.0 is incompatible with later versions of the program. The following instructions tell you how to save a user dictionary from version 2.0.

Save a Previous User Dictionary

1. Open your older version of OmniPage.
2. Choose *Edit User Dictionary...* in the Settings menu.
   The Select File dialog box appears.
3. Select *user.ud* and click *OK*.
   The Edit User Dictionary dialog box appears.
4. Click *Export...*.
   The Export to Text File dialog box appears.
5. Save the dictionary as a text file in a different directory.

Import a Saved Dictionary

1. Install and open OmniPage Pro (see next section).
2. Choose *Edit User Dictionary...* in the Settings menu.
   The Select File dialog box appears.
3. Select *user.ud* and click *OK*.

The Edit User Dictionary dialog box appears.

4 Click *Import...*.
The Import Text File dialog box appears.

5 Select the user dictionary you saved as a text file and click *OK*.
The information in the old user dictionary is added to the new user dictionary.
See “Edit User Dictionary” on page 149 for more information.

Setting up a Windows Swap File

Although 8MB is the minimum amount of memory required, OmniPage can perform faster with more memory. 12–16MB RAM is recommended for optimal performance. Set up a permanent Windows swap file with a minimum of 8MB of free, contiguous disk space to improve disk speed.

A swap file acts as virtual memory. Free disk space set aside as a swap file is used as if it were additional memory. This lets you run more programs than you could with memory alone.

The disk space used for a swap file is different than the disk space needed for temporary storage while you are working on a file. Be sure to allocate enough free disk space for both a swap file and temporary storage.

Windows 3.1 automatically creates a swap file at setup. You can change its size. You may need to *defragment* the disk first to make sure free disk space is in one empty block instead of fragmented into smaller portions.

Use a program such as Norton Utilities to defragment a hard disk. You could also exit Windows and type `defrag` at the DOS prompt if you have version 6.0 or later of DOS. For more information about swap files, see the Optimizing Windows chapter in your Windows User’s Guide.

To set up a Windows swap file (virtual memory):

1 Start Windows in Enhanced mode by typing `win /3`.

2 Double-click the Control Panel icon in the Main window of the Program Manager.

3 Double-click the 386 Enhanced icon to open the 386 Enhanced dialog.

4 Click the *Virtual Memory* button to open the Virtual Memory dialog.

This dialog displays the location, size, and type of swap file. The swap file should be at least 8192KB.

5 Click the *Change* button to expand the dialog box.

6 Select a new drive in the *Drive* list if you want to locate the swap file some place other than the default drive.

For example, you can store the swap file on a second hard disk that is faster or larger than the default. If you cannot find a drive with at least 8192KB of free space, try deleting some files and optimizing the disk again.

Create your swap file in an uncompressed drive. If you use DoubleSpace or another disk compression method, consult its documentation regarding swap files.

7 Select *Permanent* in the *Type* list.

8 Type `8192` or greater in the *New Size* edit box and select *Use 32-Bit Disk Access* if it is available.

9 Click *OK* in the Virtual Memory dialog box and click *Yes* to verify changes to virtual memory.

10 Restart Windows.

Installing the Software

Close all applications — including screen savers and mail applications — to free up memory before installing OmniPage Pro.

1 Start Windows and open the Program Manager window.

2 Insert OmniPage Pro disk #1 in drive `a:` (or `b:`) of your computer.

3 Choose *Run* in the Program Manager File menu.

The Run dialog appears.

4 Type `a:\setup` (or `b:\setup`) in the *Command Line* text box and click *OK*.

A dialog box prompts you to choose where to install OmniPage. The default directory is `c:\omnipro`.

5 Click *Continue* to start installation or type your desired location and then click *Continue.*

A dialog box warns that all executable files will be deleted from your current `omnipro` directory if you have one.

- Click *Continue* to continue.
- Click *Back* if you want to return to the previous dialog box.

A progress meter appears if you click *Continue.*

6 Insert the other installation disks as prompted.

A dialog box prompts you to install a scanner driver when installation of OmniPage is complete.

7 Do one of the following:

- Proceed to "Scan Manager Installation" on page 8 if you are using a scanner.
- Click *Exit* to finish your OmniPage installation if you are not using a scanner.

An OmniPage Pro icon is added to the Caere Applications program group. Restart Windows. You cannot use the Direct Input feature until you restart Windows. See Chapter 5, Direct Input, for information.

Scan Manager Installation

You must install the Scan Manager if you plan to use a scanner with OmniPage. Be sure your scanner is connected, compatible with your system, and runs with the software provided by the manufacturer *before* you install the Scan Manager.

You are prompted to install the Scan Manager following OmniPage installation. You will use the Scan Manager to install scanner drivers and select a default scanner.

Follow instructions in "Installing the Software" on page 7 to install OmniPage first if you have not done so.

1 Insert the disk labeled Scan Manager disk as prompted at the end of OmniPage installation and click *Continue.*

A dialog box informs you that the program will create certain directories. It lists the files that will be copied to these directories.

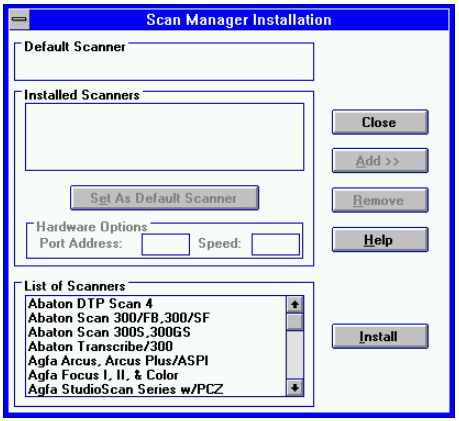
2 Click *Continue.*

The Scan Manager installs and a dialog box asks if you want to install a scanner driver now.

3 Click *Yes* to continue or *No* to exit.

You can install a scanner driver anytime after Scan Manager installation if you click *No*. See the next section for instructions.

4 Click *Yes* to install a scanner driver now.
The Scan Manager Installation dialog box appears.



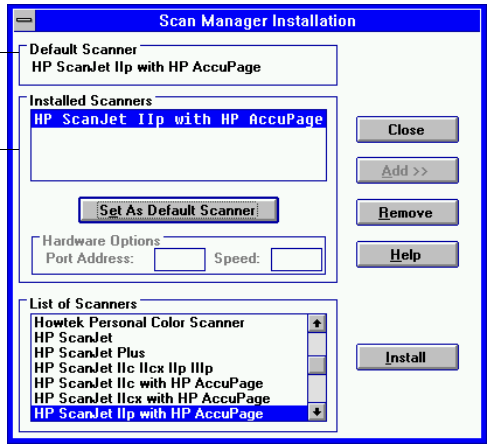
5 Locate and select your scanner in the *List of Scanners* list box.

6 Click *Install.*
The scanner appears in the *Installed Scanners* list box.

7 Click *Set As Default Scanner.*
The scanner appears in the *Default Scanner* list box.

The default scanner is used when you scan in OmniPage.

Installed scanner drivers appear here. You can install more than one.



You can install more than one scanner driver but only one can be the default scanner. Repeat steps 5 and 6 to install more drivers.

8 Click *Close* when you are done.

9 Restart Windows.

You cannot use the Direct Input feature until you restart Windows. See Chapter 5, Direct Input, for information.


An OmniPage Pro and a Scan Manager icon are added to the Caere Applications program group.

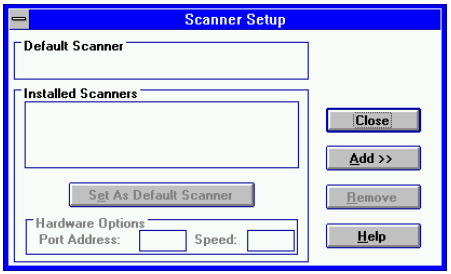
Selecting Your Scanner After Scan Manager Installation

You can install a scanner driver anytime after Scan Manager installation.

1 Exit OmniPage if it is running.

2 Double-click the Scan Manager icon in the Caere Applications program group.

The Scanner Setup dialog box appears.



3 Click *Add*>>.

4 Insert the Scan Manager disk when prompted.

The dialog box expands to show a list of available scanner drivers.

5 Follow steps 5 through 8 in the previous section.

Changing the Default Scanner Selection

You can change your default scanner selection anytime.

1 Exit OmniPage if it is running.

2 Double-click the Scan Manager icon in the Caere Applications program group.

The Scanner Setup dialog box appears.

3 Skip to step 8 if you just want to change your default scanner and do not need to install a new scanner driver. Proceed to step 4 to add a new scanner to the *Installed Scanners* list box.

4 Click *Add*>>.

5 Insert the Scan Manager disk when prompted.
The dialog box expands to show a list of available scanner drivers.

6 Select a scanner in the *List of Scanners* list box.

7 Click *Install.*
The scanner appears in the *Installed Scanners* list box.

8 Select the scanner in the *Installed Scanners* list box that you want to be the default scanner.

9 Click *Set As Default Scanner.*
The scanner appears in the *Default Scanner* list box.

10 Click *Close.*

Make sure the scanner you selected is already attached to your computer, turned on, and working when you next launch OmniPage.

Starting OmniPage Pro

To start OmniPage:

1 Double-click the OmniPage Pro icon in the Caere Applications program group.
The first time you launch OmniPage, the User Information dialog box appears.

2 Type your name in the *Licensee* text box.

3 Type your company name in the *Company* text box if you are with a company; otherwise, leave it blank.

4 Click *OK.*

The Product Registration dialog appears the first time you launch OmniPage.

5 See the next section for instructions on how to register OmniPage.

A scanner message may appear when you close this dialog. See "Scanner Message on Launch" on page 227.

Registering OmniPage

You can use OmniPage for 25 sessions without registering it. A Register menu appears in OmniPage until you register your copy. See "The Register Menu" on page 151 for information. After 25 sessions, the Registration dialog appears when you launch OmniPage, but the program exits if you click *Cancel*.

Registering your copy of OmniPage entitles you to technical support, notification of special offers and upgrades, and the lowest price offered on the next OmniPage upgrade.

1 Click the *Call* drop-down list to find the number you should call from your country.

2 Call the number and ask for a registration number.

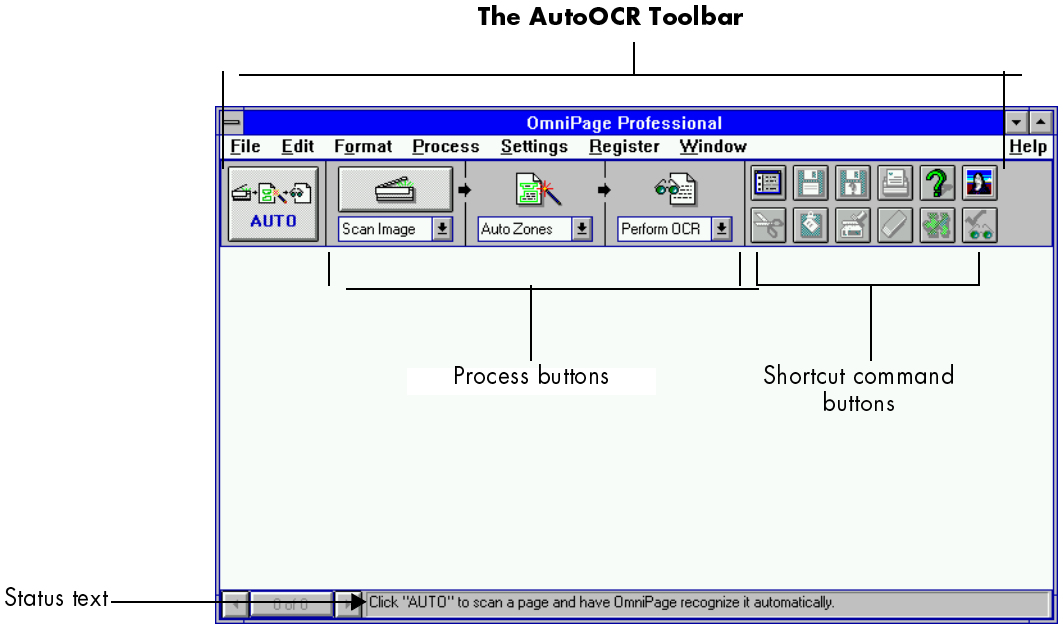
You will be asked to provide some information and you will be assigned a registration number.

3 Enter the number in the *Registration Number* text box and click *OK*.

You are now a registered user of OmniPage.

The OmniPage Window

The OmniPage window and AutoOCR™ toolbar appear after the Registration dialog box closes.



Refer to Chapter 2, Tutorials, for an overview of OmniPage tools and recognition techniques. The tutorials begin with basic scanning and an overview of the OmniPage window and move on to more advanced exercises.

Chapter 2

Tutorials

This chapter contains eight tutorials. The tutorials take you through practical exercises for everyday documents such as multi-column pages and spreadsheets. They also cover more advanced concepts such as how to use manual zoning and using deferred page recognition to maximize efficiency.

The following tutorials are in this chapter:

- Tutorial 1 — Introduction to OmniPage
- Tutorial 2 — Basic Text Recognition
- Tutorial 3 — Working With Graphics
- Tutorial 4 — Evaluating a Page
- Tutorial 5 — Scanning a Single Column or Table
- Tutorial 6 — Train OCR
- Tutorial 7 — Deferring OCR
- Tutorial 8 — Using Direct Input

See the Table of Contents for a list of exercises within each tutorial.

Be sure your scanner is attached, turned on, and compatible with your system. Test the scanner with the manufacturer's software to ensure that it works properly before using it with OmniPage.

Tutorial 1 — Introduction to OmniPage

This tutorial gives you a brief introduction to OmniPage. It contains the following sections:

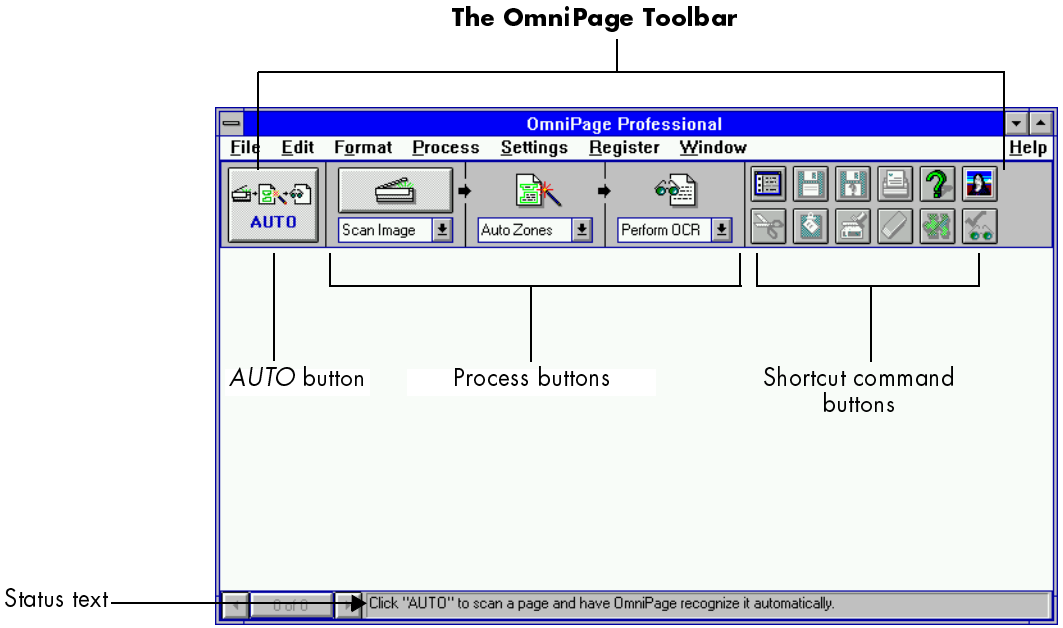
- Launch OmniPage
- What is Optical Character Recognition (OCR)?
- The OCR Process
- Scan the Quick Scan Page Sample
- Settings Panel Overview

You will use the Quick Scan Page sample in this tutorial.

Launch OmniPage

Double-click the OmniPage icon in the Caere Applications program group to launch OmniPage.

The OmniPage window opens.



The toolbar contains an *AUTO* button, three large process buttons, and the smaller shortcut command buttons.

Status text appears at the bottom of the window. It tells you what you can do next or what is taking place at the moment in the OCR process.

What is Optical Character Recognition (OCR)?

Optical character recognition (OCR) is the process of converting an image file to editable text or graphics. An image is an electronic picture of text and/or graphics. You acquire an image in two ways:

- By scanning a hard-copy document
- By loading an image file such as a TIFF or PCX file (for example, a received fax file can be saved to an image-file format and recognized in OmniPage)

The image you scan or load is at first just a "picture" to your computer even if it contains text. You can see the text but you cannot edit it. You need to perform OCR to turn the image into individual characters.

During OCR, OmniPage looks for and defines characters on the image to produce editable text. You can export the recognized text from OmniPage to a variety of word-processing, page layout, and spreadsheet programs. OCR is also referred to as *text* or *page recognition*, or just *recognition*.

The OCR Process

OCR is a three-step process: acquiring an image, zoning the image, and recognizing the image. Use the process buttons in the toolbar (or corresponding commands in the Process menu) to set up the OCR process.

The Process Buttons

Each of the following three buttons represents one step in the optical character recognition (OCR) process.

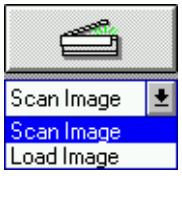
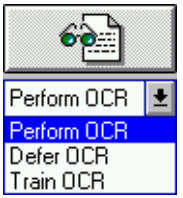


Image button

Zone button



OCR button

Using the Process Buttons

The OCR process offers choices at each step:

1 Loading an image into OmniPage

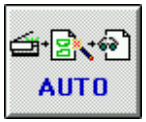
- Select *Scan Image* in the Image button's drop-down list to scan in a hard-copy document with a scanner.
- Select *Load Image* to import a graphic-format file such as TIFF or PCX.

2 Setting recognition zones

- Select *Auto Zones* in the Zone button's drop-down list to have OmniPage define the page areas to be recognized.
- Select *Manual Zones* to draw the zones yourself.

3 Performing OCR on the zoned page areas

- Select *Perform OCR* in the OCR button's drop-down list to perform optical character recognition on a zoned page.
- Select *Defer OCR* to perform OCR later.
- Select *Train OCR* to teach OmniPage to recognize special characters before OCR.



You can either click each process button individually to activate its process or click the *AUTO* button to activate the buttons automatically depending on what is selected in the drop-down lists.

In the following tutorial exercises, you will select a processing option for each stage of OCR before you load an image or scan a document.

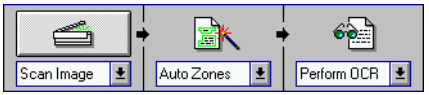
Scan the Quick Scan Page Sample

You will scan the Quick Scan Page Sample in this exercise for a quick introduction to the OCR process.

Select the Settings

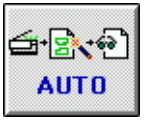
1 Click the drop-down list under each process button and select these options:

- *Scan Image*
- *Auto Zones*
- *Perform OCR*



Scan the Page

1 Place the Quick Scan Page Sample in your scanner making sure it is aligned correctly.



2 Click the *AUTO* button or choose *Auto* in the Process menu.

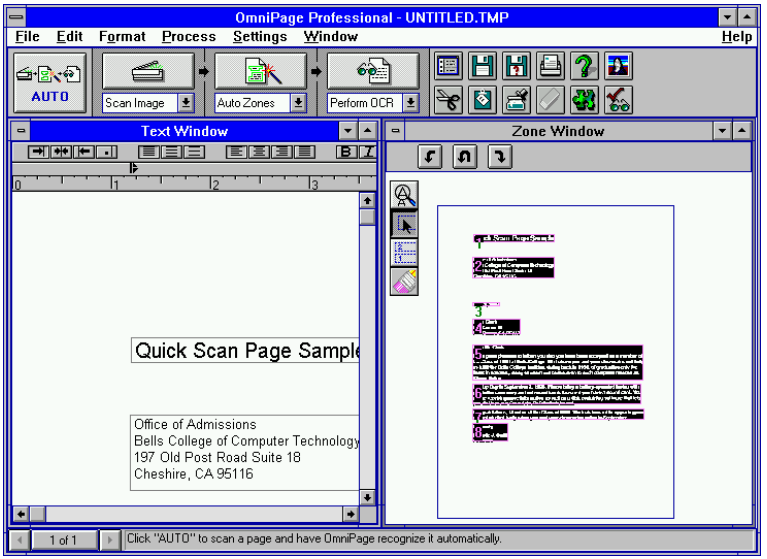
- OmniPage scans the page and opens it in the zone window.
- Automatically drawn zones appear on the image to show how text will be ordered.

- OmniPage makes three recognition passes over the page: cyan, light blue, and dark blue.

Each of these three stages is discussed in more detail in later tutorials.

OmniPage opens the recognized page in a maximized text window.

3 Choose *Tile Vertical* in the Windows menu so that you can see both the zone and text windows.



The zone window shows the scanned image of the page. Note that although you can see the text, you cannot select words or letters, or edit the text in any way.

The text window shows the recognized, editable text.

4 Double-click the word *Computer* in the text window.

The Verification window opens to show the image of the word as it was scanned originally.

You can retype the highlighted word if necessary while the Verification window is still open. This is a quick way to edit text without using the spell checker.

5 Click anywhere in the text window to close the Verification window.

6 Choose *Close Document...* in the File menu to close the Quick Scan Page sample.

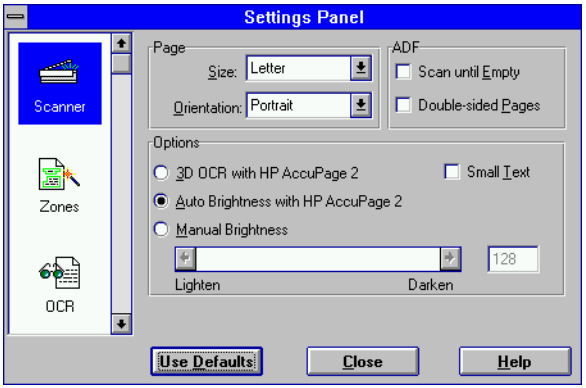
7 Click *No* in the dialog box that asks if you want to save changes.

You will edit documents and save them in later tutorials.

Settings Panel Overview

Use the Settings Panel to customize the OCR process for particular pages. The page you just scanned had a simple page layout with crisp black text on a clean white background. Most settings work well with this type of page. You will customize the Settings Panel options in later tutorials.

1 Click the Settings Panel button in the toolbar or choose *Settings Panel...* in the Settings menu.
The Settings Panel appears.



There are seven panels in the Settings Panel dialog box: Scanner, Zones, OCR, Fonts, Spelling, Direct Input, and Preferences.

2 Click each icon in turn to view its options.
Use the scroll box to access and select icons below the OCR icon.

3 After you click the Preferences icon, click *Close*.

4 Position the mouse pointer over the Image button in the toolbar and click the *right* mouse button.

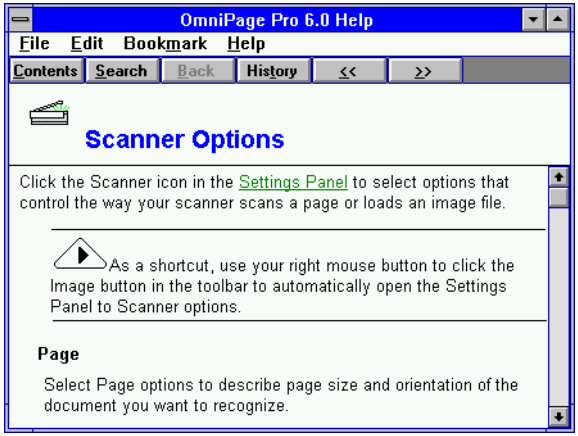
5 The Settings Panel opens to the Scanner options.

You can also click with the right mouse button on the Zone and OCR process buttons when they are active to open the Settings Panel to the corresponding settings. These two buttons are active after a document has been loaded or scanned.

You would set Scanner options before scanning a page. Your Scanner settings panel options may look different than those pictured above, depending on your scanner.

6 Click *Help.*

The online Help program opens to Scanner Options.



This section of the Help program gives information on all the Scanner settings panel options. You can click the *Help* button in each settings panel to open its corresponding Help section.

7 Choose *Exit* in the File menu to close the online Help.

8 Click *Close* to close the Settings Panel.

See Chapter 4, The Settings Panel, for detailed information on all settings. The next tutorial introduces you to more scanning concepts.

Tutorial 2 — Basic Text Recognition

This tutorial takes you through basic scanning, zoning, and OCR exercises with OmniPage. It contains the following exercises:

- Scanning With the Default Settings
- Change a Document's Fonts During OCR
- Ignore All Formatting
- True Page Recognition
- Deselect Retain Graphics
- Save a Settings File
- Load an Image File

You will use the True Page sample in these tutorials.

Save the files as directed during the exercises so you can use them in later exercises.

Scanning With the Default Settings

You will select the default settings in this exercise, observe the OCR process, use the *Check Recognition* command to correct any recognition errors, and save the recognized document in two different file formats.

1 Click the drop-down list under each process button and select these options:

- *Scan Image*
- *Auto Zones*
- *Perform OCR*

2 Click the Settings Panel button or choose *Settings Panel...* in the Settings menu.

3 Click *Use Defaults.*

4 Click *OK* in the dialog box that asks if you are sure.

5 Click the Scanner icon.

Auto Brightness with AnyPage/HP AccuPage 2 is the default. (HP stands for Hewlett-Packard.)

This setting works well with most types of pages. The default is *Manual Brightness* if you have a black-and-white scanner.

6 Click the Zones icon.

The default setting is *Multiple Columns.*

The True Page sample has multiple columns so this is the setting you want. Use this option for pages such as newsletters, data sheets, and magazines.

7 Click the OCR icon.

Retain Font and Paragraph Formatting is the default setting under the section *Output Formatting Options.*

This setting preserves paragraph order and formatting (centered or left-aligned), and font style (serif and sans serif) and formatting (bold, point size, etc.) during OCR.

8 Click *Close.*

You can leave the Settings Panel open if you have room on your screen. This is useful if you need to change the settings frequently.

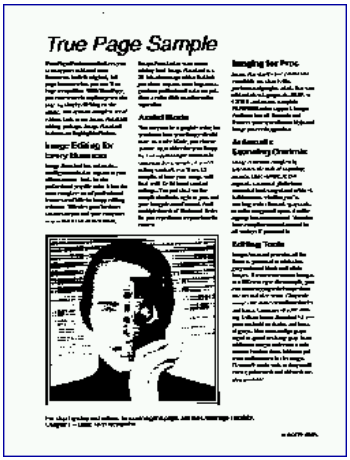
Scan the Page

You will click the process buttons individually in this exercise to observe each stage of the recognition process.

1 Place the True Page Sample in your scanner making sure the page is aligned correctly.

2 Click the Image button.

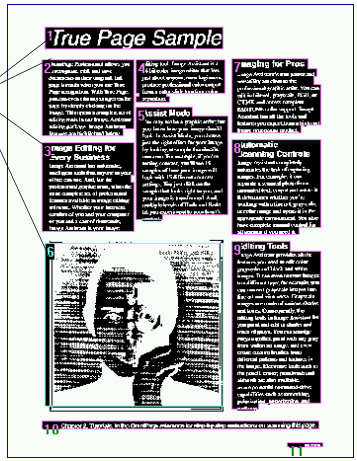
OmniPage scans the page and opens the image in the zone window.



3 Click the Zone button.

OmniPage determines column flow and automatically draws zones. This shows how text and graphics will be ordered during OCR.

Numbered zones indicate recognition order.

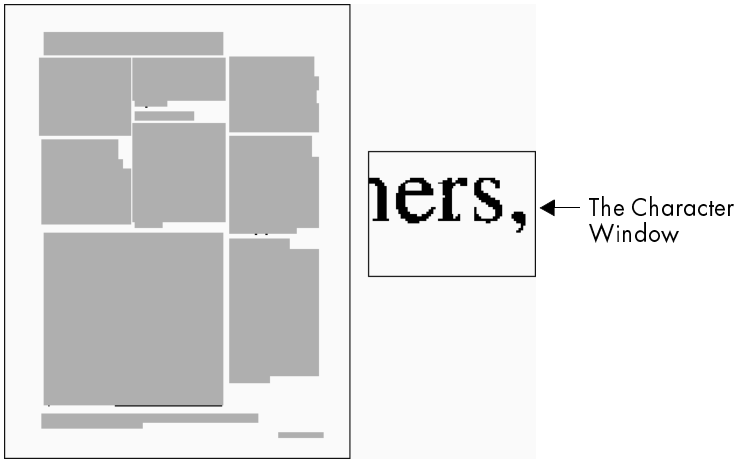


Your zones may be different depending on whether you are using AnyPage, HP AccuPage, or Manual Brightness.

4 Click the OCR button.

OmniPage performs three OCR passes over the document: a cyan pass for initial recognition; a blue pass as text is analyzed and corrected; and, a dark blue pass for final recognition.

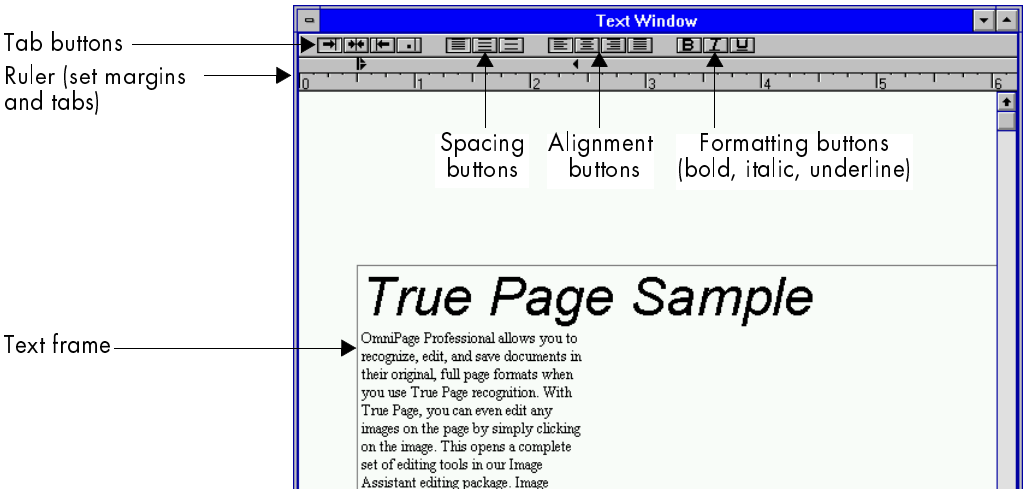
The Character window displays characters as OCR takes place.



The recognized document opens in a new maximized text window. See the next section for an overview of the text window and its editing tools.

The Text Window

The document's font and paragraph formatting are retained but page layout is not. Text is displayed in one column with the graphic at the end.



If the text is not ordered correctly, you may have misaligned the page in your scanner. Realign the page and try scanning again.

1 Choose *Tile Vertical* or *Tile Horizontal* in the Windows menu.
The text and zone windows tile for easy viewing.

2 Compare the recognized document in the text window to the scanned image in the zone window.
OmniPage highlights any words it had trouble recognizing.
- Green: *suspects*, words that may not have been recognized correctly, are highlighted in green.
- Red tilde: *rejects*, or unrecognizable characters, are marked with a red tilde (~).

3 Select a word in the text.
If you double-click the word, the Verification window opens. You can still edit the word if this window is open. Click anywhere outside the Verification window to close it.

4 Click the Bold button in the text window.
The text becomes bold.

5 Experiment with the other tools in the text window to see how they affect your text.

See “The Format Menu” on page 113 for detailed information on formatting options.

The next section shows you how to correct any recognition errors and add words to a user dictionary.

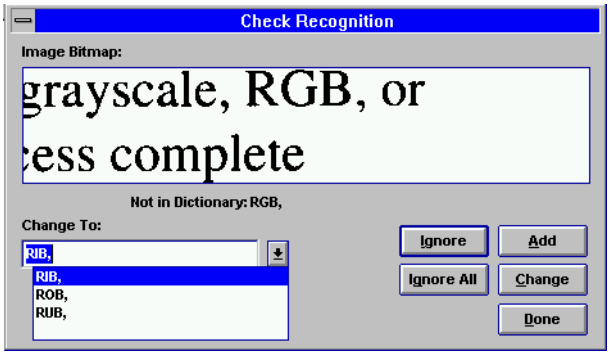
Check Recognition

The True Page Sample has black, crisp text on a clean white background and so should have few, if any, recognition errors. Check Recognition, however, also allows you to add words to your user dictionary as well as correct recognition errors.

1 Click the text window to make it active if it is not already.

2 Click the Check Recognition button or choose *Check Recognition...* in the Edit menu.

The Check Recognition window appears. It displays the image and text of any questionable or unrecognizable word.



3 Correct any errors in the text.

If the word is misspelled:

- Correct the spelling in the *Change To* edit box and click *Change.*

  OmniPage may list one or more suggestions in the *Change To* drop-down list. The first word in the list is the word as OmniPage recognized it. Select a word in the list and click *Change* to replace the word in the text. Alternatively, type the proper word in the *Change To* edit box.

If the word is correct:

- Click *Add* to add the word to the User Dictionary. The word will still be flagged if it is a suspect (green) word and it occurs again.
- Click *Ignore* to ignore the currently flagged word. Other instances of the word in the document will be checked.
- Click *Ignore All* to ignore all instances of the currently flagged word in the document.

OmniPage automatically moves to the next word after you click a button.

4 Click *Done* if you want to end the spell check.

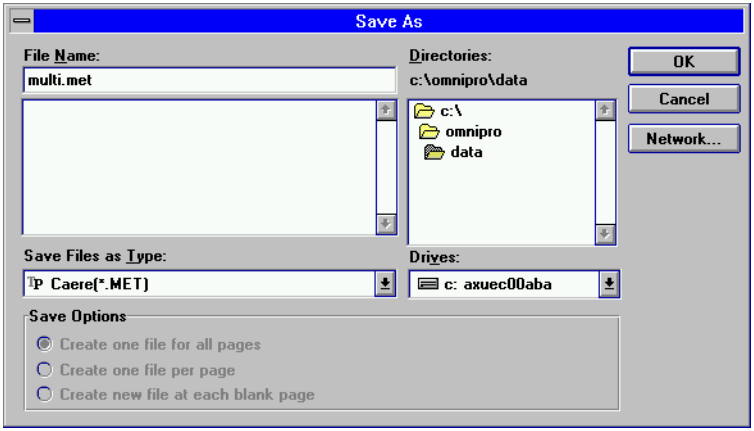
Otherwise, a dialog box informs you when the end of the document has been reached. Click *OK* in this dialog box.

Save the Document

You will save the document as a Caere Document (a special OmniPage format), reopen it, and save it as a word-processing file.

Save as a Caere Document

1 Click the Save As... button or choose *Save As...* in the File menu. The Save As dialog box opens.



2 Select *Caere[*.MET]* in the *Save File as Type* drop-down list. The `data` directory is the default location, but you can choose another location if you wish.

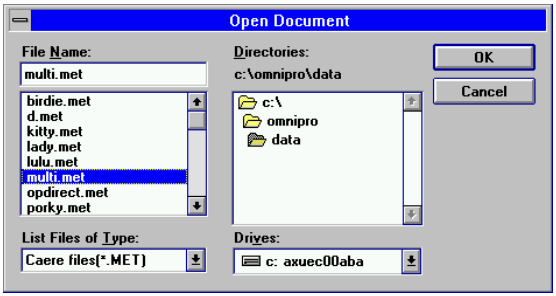
3 Type `multi.met` in the *File Name* edit box.

4 Click *OK*.

5 Choose *Close Document* in the File menu.

Reopen the Document

1 Choose *Open Document...* in the File menu.
The Open dialog box appears.



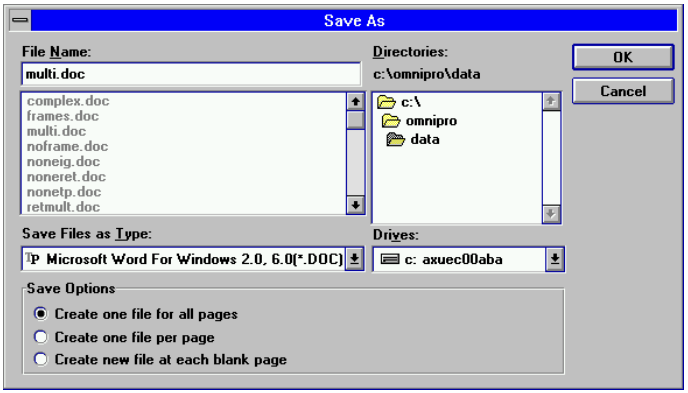
2 Select `Caere files[*.MET]` in the *List Files of Type* drop-down list if it is not selected already.

3 Locate and open the file `multi.met`.
The text window opens maximized.

OmniPage opens only Caere Documents and image files. A Caere Document can contain both text and zone window information from a recognized document. (An image file contains only an image.) You can save a Caere Document to multiple file formats. You can also rezone or re-recognize it to save the time of rescanning the original document.

Save as a Word-Processing File

1 Click the Save As... button or choose *Save As...* in the File menu.
The Save As dialog box appears.



2 Select a word-processing application file type in the *Save Files as Type* list box, such as Microsoft Word for Windows.

Type a new name for the file in the *File Name* text box if you like.

3 Click *OK*.

4 Leave the document open for the next exercise.

Change a Document's Fonts During OCR

In the previous exercise, OmniPage retained font formatting but *mapped* the fonts to ones preselected in the Fonts settings panel. You can change the fonts and point sizes assigned to your recognized document during OCR. You may want to do this to save formatting time later, either in the text window or in your target application.

You will see how font mapping works in this exercise.

Change the Font Settings

1 Choose *Open...* in the File menu to locate and open the `multi.met` file if you did not leave it open after the previous exercise.

See "Reopen the Document" on page 28 for information.

2 Click the Settings Panel button or choose *Settings Panel...* in the Settings menu.

3 Click *Use Defaults* if you have changed the settings since the last exercise.

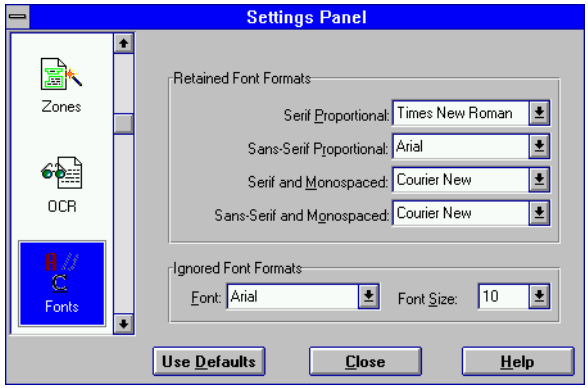
4 Click *Yes* in the dialog box that asks if you are sure.

Retain Font and Paragraph Formatting is the default OCR setting as you may recall from the last exercise.

This setting preserves paragraph order and formatting (such as centered or left-aligned), and font style (serif and sans serif) and formatting (bold, point size, etc.) during OCR.

It matches font types to the fonts selected in the *Retained Font Formats* section of the Fonts settings panel. It does not try to retain page layout.

5 Click the Fonts icon and observe the settings.



- The default *Serif Proportional* setting is *Times New Roman*. (A *seriffed* font has short lines, or *serifs*, on the ends of the strokes of a letter.)

  The body text in the True Page sample is already Times New Roman and so would not change during OCR.

- The default *Sans Serif Proportional* settings is *Arial. (*A *sans seriffed* font has no serifs.)

  The title and subtitles in the True Page sample are already Arial and so would not change during OCR.

- There are no monospaced fonts in the True Page sample so ignore these settings for now.

You can change the selection in any of the drop-down lists and the fonts in your document will change accordingly during OCR.

6 Select *Century Schoolbook* OR the font of your choice in the *Serif Proportional* drop-down list.

7 Select *Helvetica* OR the font of your choice in the *Sans Serif Proportional* drop-down list.

8 Click *Close.*

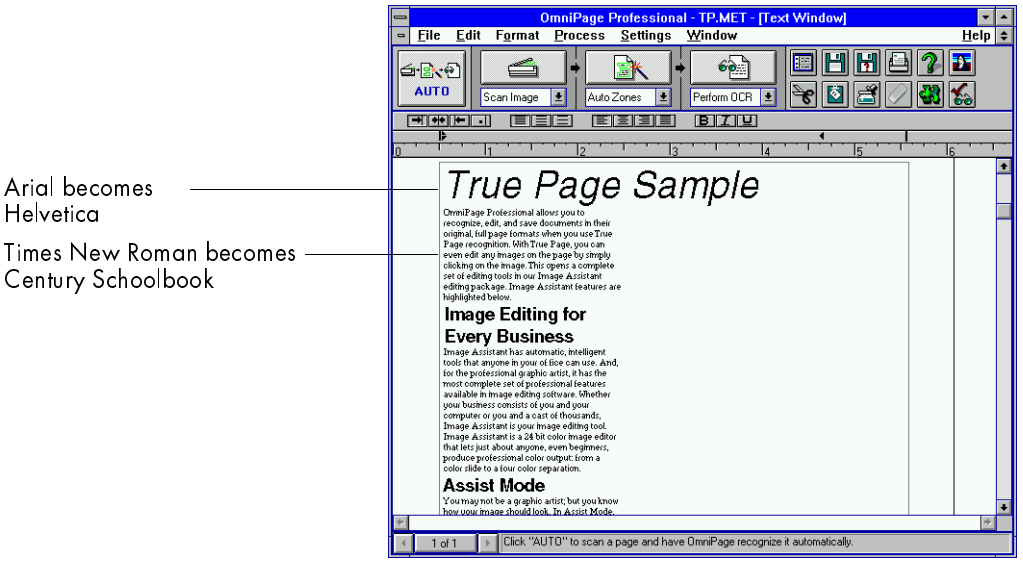
Re-recognize the Page

1 Click the OCR button.

2 Click *Yes* in the dialog box that asks if you want to replace the current text.

OmniPage re-recognizes the page and displays the recognized text in the text window.

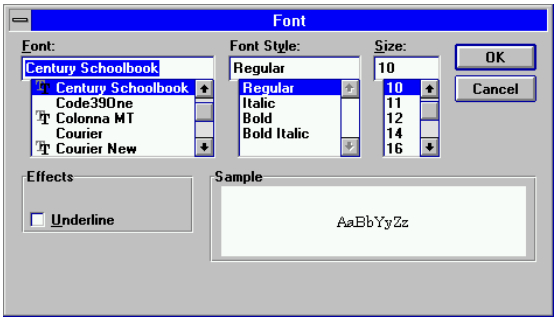
Font and paragraph formatting are retained but page layout is not. Text is displayed in one column with the graphic at the end. The fonts match the selections in the Fonts settings panel.



3 Click in the body of the text.

4 Choose *Font...* in the Format menu.

The Font dialog box appears.



5 Verify that the font display matches the font you selected in the *Serif Proportional* drop-down list in the Fonts settings panel.

6 Leave the document open for the next exercise.

See "Retain Font and Paragraph Formatting" on page 172 for detailed information on the Fonts settings panel options.

Ignore All Formatting

You may decide you do not need any formatting at all, just the recognized text itself. You will use the *Ignore All Formatting* OCR option to strip away font and paragraph formatting during recognition and assign one font and point size to the recognized text.

This option speeds the OCR process. It is useful if you want to export just text that either needs no particular formatting or that you want to format yourself in your target application.

1 Choose *Open...* in the File menu to locate and open the `multi.met` file if you did not leave it open after the previous exercise.

See "Reopen the Document" on page 28 for information.

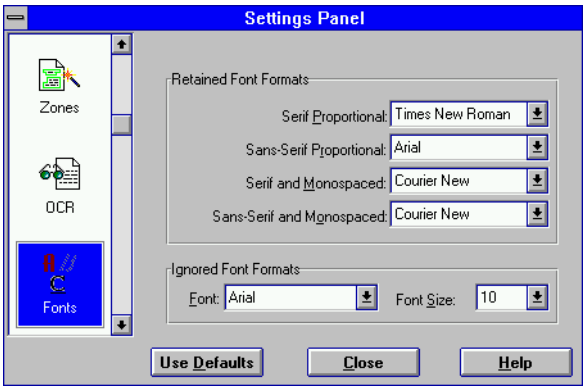
2 Click with your *right* mouse button on the OCR button to open the Settings Panel to the OCR settings.

3 Select *Ignore All Formatting*.

OmniPage will maintain paragraph order but not formatting. It will ignore font types (serif and sans serif) and any formatting (bold, point size, etc.) when it recognizes the document. You can choose a single font type and size for all the recognized text in the *Ignored Font Formats* section of the Fonts settings panel.

4 Click the Fonts icon.

The Fonts options appear.



The default setting under *Ignored Font Formats* is Arial 10-point. All recognized characters will be formatted as plain, Arial, 10-point text. You can choose a different font and point size in the drop-down lists if you like.

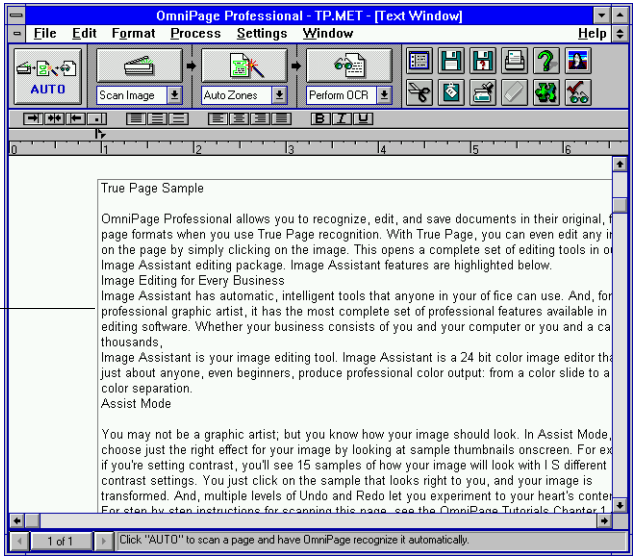
5 Click *Close*.

6 Click the OCR button.

7 Click *Yes* in the dialog box that asks if you want to replace the text.

OmniPage re-recognizes the page and displays the recognized text in the text window.

All text is now 10-point Arial.



Formatting has been discarded and all text is Arial 10-point (or whichever font and point size you chose). The text is displayed in one column in order of recognition with the graphic at the end.

8 Leave the document open for the next exercise.

True Page Recognition

You may want to scan a document and retain not only font and paragraph formatting, but also as much page layout as possible. You can retain page layout by using the *True Page - Retain All Page Formatting* OCR option.

You will re-recognize the True Sample with the True Page OCR option, work with frames in the text window, and deselect the Retain Graphics option to observe what effect this has on True Page recognition.

1 Choose *Open...* in the File menu to locate and open the `multi.met` file if you did not leave it open after the previous exercise.

See “Reopen the Document” on page 28 for information.

2 Click with your *right* mouse button on the OCR button to open the Settings Panel to the OCR settings.

3 Select *True Page - Retain All Page Formatting*.

Use this option when you want to duplicate page layout as closely as possible.

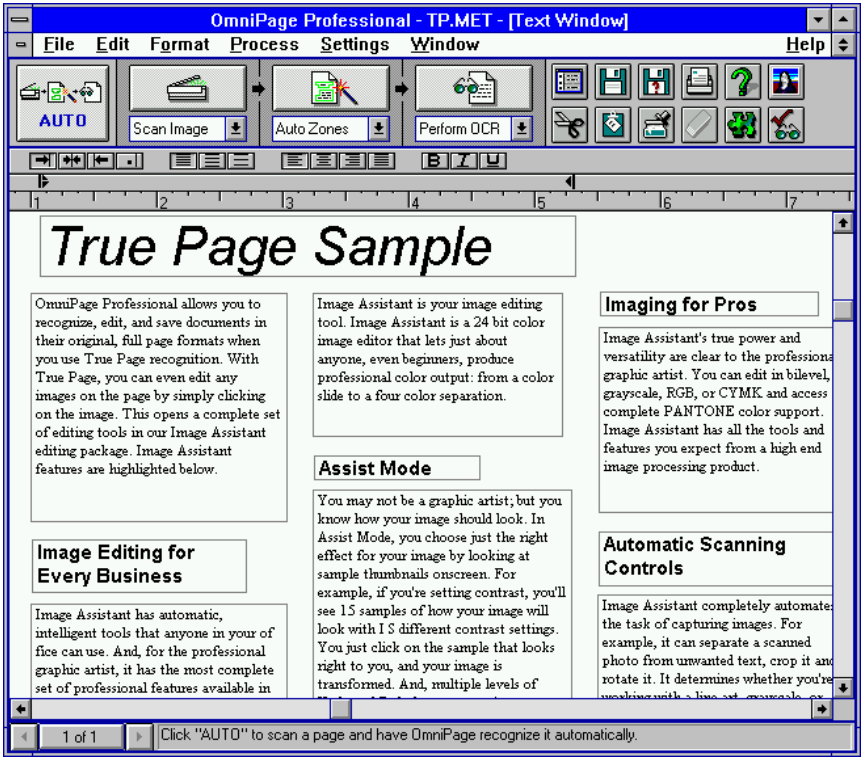
4 Click *Close*.

5 Click the OCR button.

6 Click *Yes* in the dialog box that asks if you want to replace the current text.

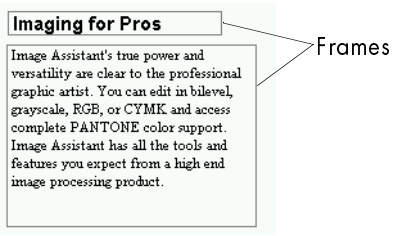
OmniPage re-recognizes the document and displays the recognized text in the text window.

The result matches the original page layout as closely as possible.



Working With Frames

Because *Multiple Columns* was the default zoning method, True Page automatically creates *frames* around recognized text and graphic zones to preserve a side-by-side column structure.



You can resize frames and move them around to modify your document's page layout. These frames are exported intact when you save your document in an appropriate file format. You will work with True Page frames in this exercise.

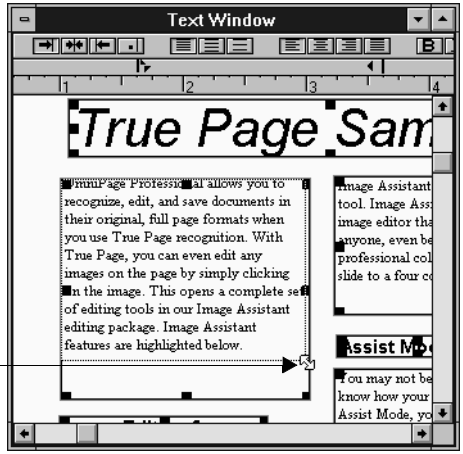
1 Click the text window to make it active if it is not already.

2 Choose *Select Recognized Zones* in the Edit menu.

You cannot select this command if the zone window is active.

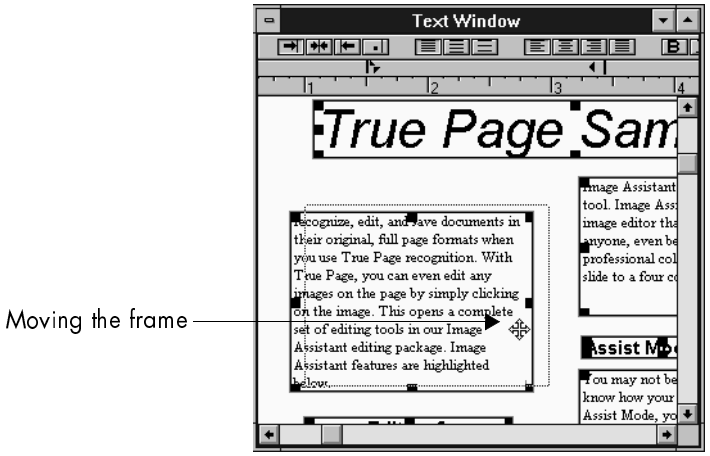
All text and graphic zones in the text window are selected. Handles appear around the text zones.

3 Hold your cursor over a frame handle in a text zone so that it turns into a two-way arrow.



4 Hold down the mouse button and drag to resize the frame.

5 Place your cursor inside a text zone so that it turns into a four-way arrow.



6 Hold down the mouse button and drag the zone to any location on the page.

7 Choose *Select Recognized Zone*s in the Edit menu again.

All frames are deselected. A check mark in front of the command indicates that the command is active. The check mark disappears when you reselect the command.

8 Place your cursor inside a frame.

9 Hold down the Alt key, and click the right mouse button.

This selects an individual frame.

10 Repeat the Alt-right-mouse-button click to deselect the frame.

11 Leave the document open for the next exercise.

Deselect Retain Graphics

You may want to retain page layout but not graphics during page recognition. Not retaining graphics speeds recognition because OmniPage can skip over those zones. You will re-recognize the document you scanned in the previous exercise but not retain the graphic.

1 Choose *Open...* in the File menu to locate and open the `multi.met` file if you did not leave it open after the previous exercise.

See “Reopen the Document” on page 28 for information.

2 Click with your *right* mouse button on the OCR button to open the Settings Panel to the OCR settings.

3 Deselect *Retain Graphics*.

4 Select *True Page - Retain All Page Formatting* if it is not selected already.

5 Click *Close.*

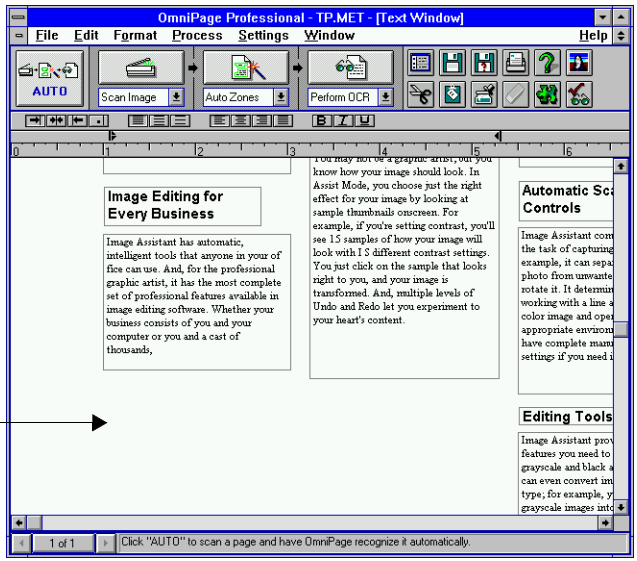
6 Click the OCR button.

7 Click *Yes* in the dialog box that asks if you want to replace the text.

OmniPage re-recognizes the page.

The text appears in the same format as before, but has an empty space where the graphic was originally.

Empty space where graphic was



8 Choose *Close Document* in the File menu.

9 Click *No* in the dialog box that asks if you want to save changes.

Save a Settings File

You may find that you use the same Settings Panel options often. You can save these settings as a file and load the file before scanning or loading an image file. This saves you the time of opening the Settings Panel and resetting the options you need.

Save the Settings

1 Click the Settings Panel button or choose *Settings Panel...* in the Settings menu.

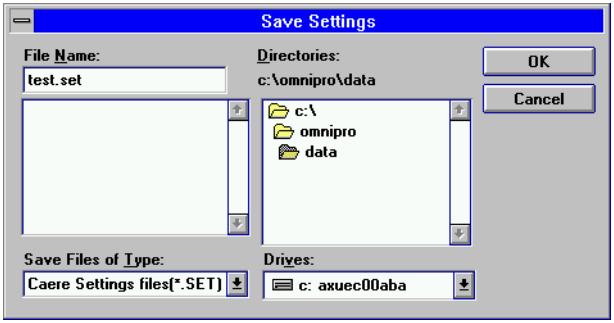
2 Select the following options in each settings panel:

- Scanner: *Manual Brightness*
- Zones: *One Zone*
- OCR: *Ignore All Formatting*

Note that none of these settings is a default setting.

3 Leave the Settings Panel open.

4 Choose *Save Settings...* in the File menu.
The Save Settings dialog box appears.



- `Caere Settings files[*.SET]` is the only selection in the *Save Files of Type* list box.
- The `data` directory is the default location but you can choose another if you like.

5 Type the name `test.set` in the *File Name* text box.

6 Click *OK*.

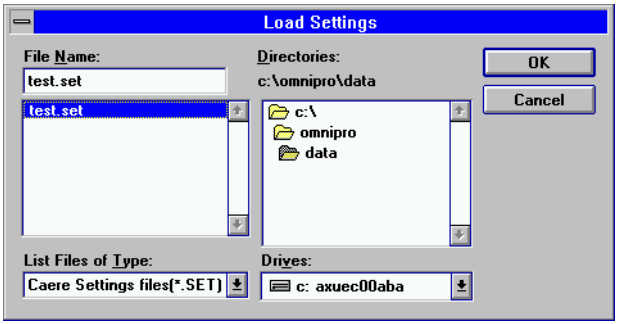
Load the Settings

1 In the Settings Panel, click *Use Defaults*.

2 Click *Yes* in the dialog box that asks if you are sure.

3 Choose *Load Settings...* in the File menu.

The Load Settings dialog box appears.



4 Locate and select the file `test.set`.

5 Click *OK*.

The Settings Panel settings change to match the settings file you just loaded.

6 Click the Scanner, Zone, and OCR icons to verify that their respective settings have changed.

7 Choose *Close Document* in the File menu.

8 Click *No* in the dialog box that asks if you want to save changes.

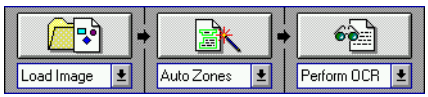
Load an Image File

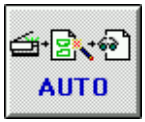
OmniPage can load, zone, and recognize TIFF and PCX files in the same way it does scanned documents. You will load an image file in this exercise and experiment with font settings. See “Supported Input File Formats” on page 239 for a complete list of supported file types you can load.

Load a Single Image File

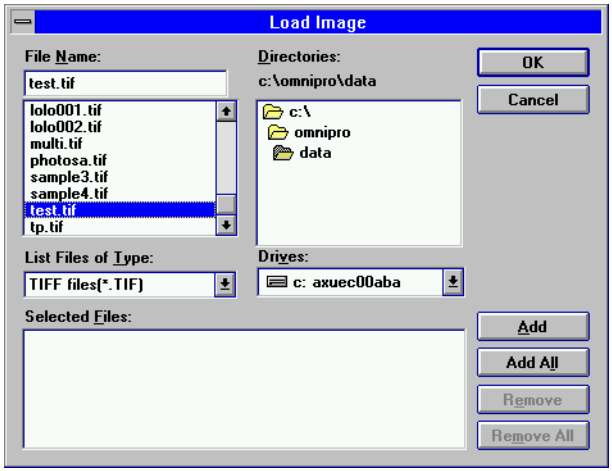
1 Click the drop-down lists under the process buttons and select:

- *Load Image*
- *Auto Zones*
- *Perform OCR*





2 Click the *AUTO* button.
The Load Image dialog box appears.



3 Select `TIFF files[*.TIF]` in the *List Files of Type* drop-down list.

4 Locate and select the `test.tif` file.
The file was placed in the `c:\omnipro\data` directory during installation.

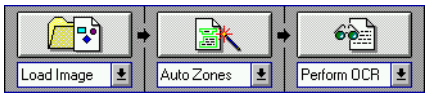
5 Click *OK*.
OmniPage loads the image file, creates automatic zones on it, performs OCR, and then displays the recognized text in the text window.

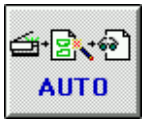
Load Multiple Image Files

You can load your own image files in this exercise if you have them. Otherwise, skip to the next tutorial. See “Supported Input File Formats” on page 239 for a list of file types you can import.

1 Click the drop-down lists under the process buttons and select:

- *Load Image*
- *Auto Zones*
- *Perform OCR*





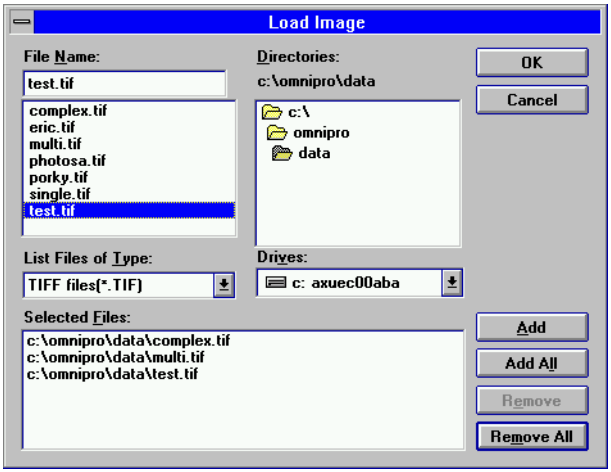
2 Click the *AUTO* button.

3 The Load Image dialog box appears.

4 Select a file format in the *List Files of Type* drop-down list.

5 Select a file to load and click *Add.*
The file appears in the *Selected Files* list box.

6 Repeat for each file you want to load.



7 Click *OK* when you have selected all the files to load.
OmniPage loads, zones, and performs recognition on the files in the order selected. The new document starts at page two if you left the document from the previous exercise open.
Each subsequent document becomes a new page in the final recognized document. Three one-page TIFF files, for example, would be merged into a three-page recognized document.

8 Choose *Close Document* in the File menu.

9 Click *No* in the dialog box that asks if you want to save changes if you do not want to save the document.
You will learn about the different save options available for multiple-page documents in the “Deferring OCR” tutorial. Or, see “Save Options” on page 92 if you want to save the document.

Tutorial 3 — Working With Graphics

OmniPage can export a scanned page or pages as one or more graphic-format files. It can also find individual graphic zones on each page and export them as graphic-format files. This tutorial contains a tutorial on how to Export a Graphic

You will use the True Page Sample in this tutorial.

Export a Graphic

You will export the graphic on the True Page sample as an individual graphic file in this exercise.

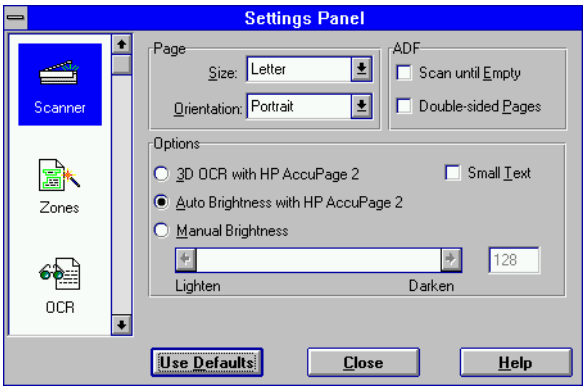
Select Settings

1 Click the drop-down list under each process button and select these options:

- *Scan Image*
- *Auto Zones*
- *Perform OCR*

2 Click the Settings Panel button in the toolbar or choose *Settings Panel...* in the Settings menu.

The Settings Panel appears.



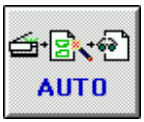
3 Click *Use Defaults*.

4 Click *Yes* in the dialog box that asks if you are sure.

5 Click *Close*.

Scan the Page

1 Place the True Page sample in your scanner making sure it is aligned correctly.

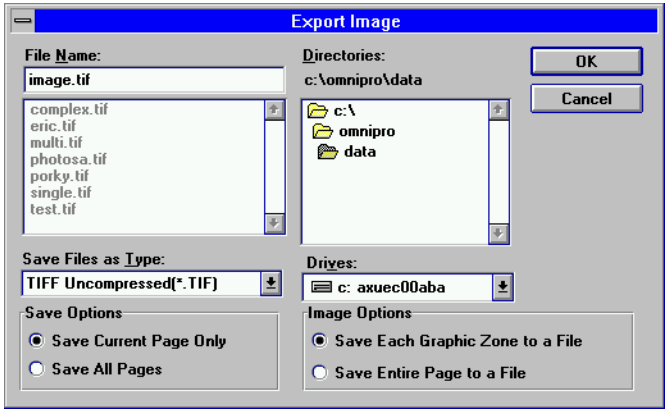


2 Click the *AUTO* button.

3 OmniPage scans, zones, and recognizes the document.

Export the Graphic Zone

1 Choose *Export Image...* in the File menu.
The Export Image dialog box opens.



2 Select *Save Current Page Only* under *Save Options*.
You only have one page, but if you had a multiple-page document open, OmniPage would save the page being viewed.

3 Select *Save Each Graphic Zone to a File* under *Image Options*.
There is one graphic zone on this page, the image of the woman. OmniPage will export just this image and none of the text.

4 Select a file format in the *Save Files as Type* drop-down list.

5 Select a location for your file.

6 Type a name for your graphic file in the *File Name* edit box.
The name you choose can have up to seven characters. OmniPage appends a letter to indicate the order of the graphic on the page.
If you had multiple graphics to export, A would indicate the first graphic, B the second and so on. Up to 26 files can be created in one directory with this method.

7 Click *OK*.

The recognized graphic zone on the page is exported in the file format you chose. You can open it in most image-editing programs.

Tutorial 4 — Evaluating a Page

A complex page may require more attention on your part for accurate OCR to take place. Tight or non-rectangular columns, text-filled or very small graphics, shading behind text, or very stylized text may be difficult for OmniPage to recognize with perfect accuracy on the first try. Sometimes you need to reprocess a page with different settings.

This tutorial illustrates some difficulties a complex page or any kind of page can present and how to correct those problems. It also gives a basic introduction to manual zoning at the same time. It contains the following exercises:

- Overcoming Recognition Difficulties
- When to Use Manual Zoning
- Manual Zones — Recognize Portions of a Page
- Manual Zones — Specify Zone Contents
- Manual Zoning — Reorder Text
- Scanning and the Brightness Setting

You will use the Complex Page sample in this tutorial.

Overcoming Recognition Difficulties

This exercise uses a fictional newsletter to illustrate some challenges you may encounter with your own scanned pages — graphics recognized as text, background interfering with text recognition, unwanted text or graphic elements on a page — and how to solve them.

Select Settings

1 Set these options in the toolbar:

- *Scan Image*
- *Auto Zones*
- *Perform OCR*

2 Click the Settings Panel button in the toolbar.

3 Select the following settings:

- Scanner: *Auto Brightness with AnyPage/HP AccuPage2*
  This is a good setting for shaded backgrounds.
- If you have a black-and-white scanner, set *Manual Brightness* to the center of the slider.
- Zones: *Multiple Columns*
  The page has multiple columns so this setting is appropriate.
- OCR: *Retain Graphics*
- You will retain the Caere logo in this exercise.
- OCR: *True Page - Retain All Page Formatting*
  This setting retains page layout and will make it easier to find various sections of the page in this exercise. You would choose *Retain Font and Paragraph Formatting* if you did not need to preserve page layout.

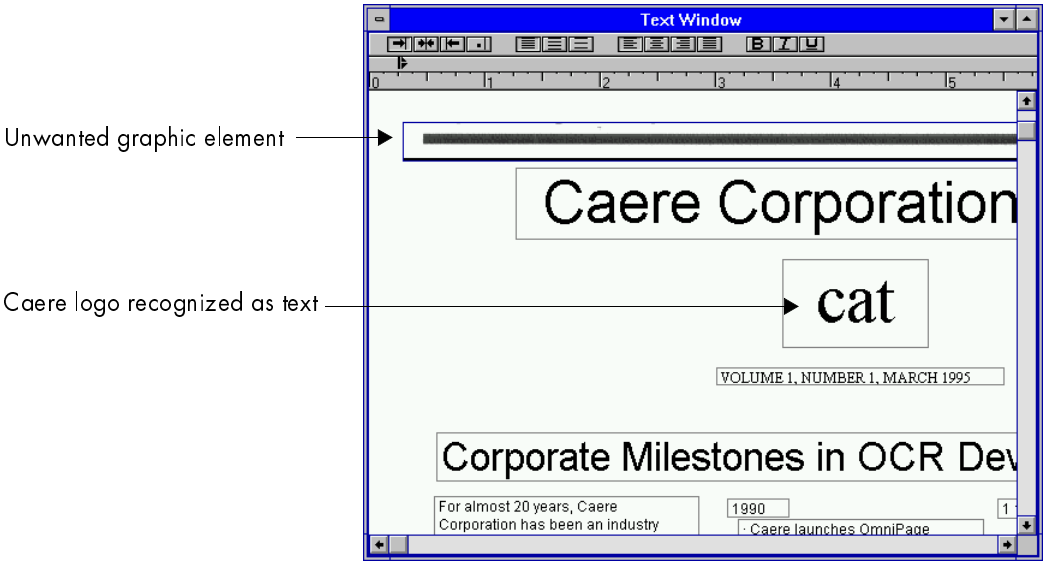
4 Click *Close*.

Scan the Page

1 Place the Complex Page Sample in your scanner making sure it is aligned correctly.

2 Click *AUTO*.

OmniPage scans, zones, and recognizes the page. The recognized page opens in the text window.



Your results may be different than those pictured above depending on your scanner. The line above the newsletter title may be not recognized at all, for example.

The Problems to be Solved

You may find some or all of the following recognition difficulties.

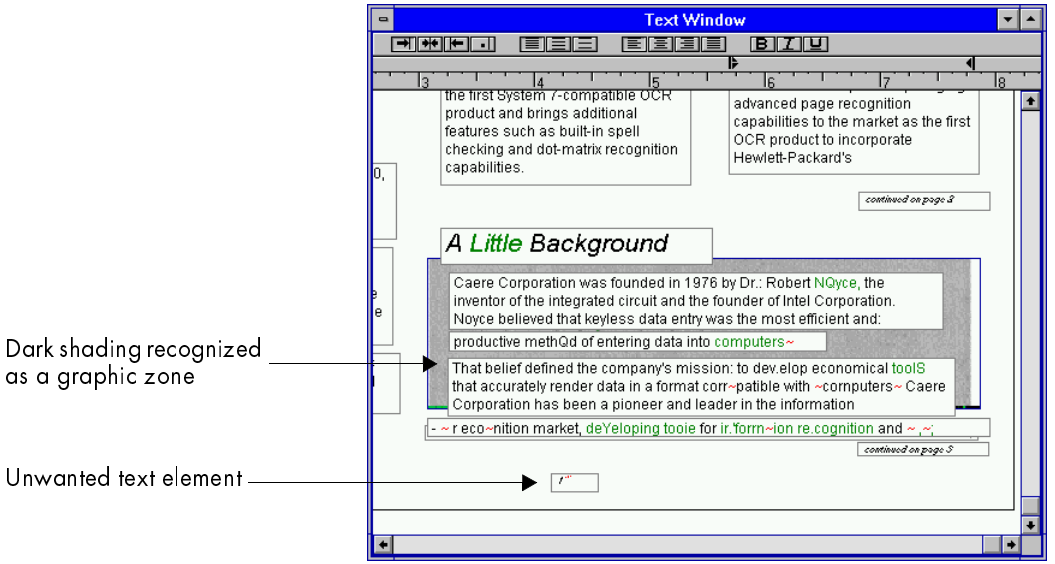
1 Note that the Caere logo was not reproduced: OmniPage tried to recognize it as text.

This is because it saw the `CA` at the beginning of the logo and assumed it was the beginning of a word such as `cat`.

OmniPage assumed `ca` was the beginning of a word.

You may have different recognition results depending on the quality of your scanner.

2 Scroll down the page to the *A Little Background* article.



OmniPage had trouble with this section because the extremely dark background could be interpreted as part of a graphic. The lack of distinct contrast also interfered with the program's ability to distinguish characters.

Depending on your scanner, you may find recognition errors and perhaps some small graphic zones here.

3 Note that OmniPage may have tried to recognize the tiny squares at the bottom of the page because they are easily confused with text. You might see tildes or other characters here.

You will not always need all the information on a page. You can choose which portions to recognize. In this exercise, you will recognize just the logo, the headlines, and the body text.

How to Solve the Problems

You will:

- Rezone the page to leave out the unwanted text and graphic elements.
- Specify a graphics zone content for the Caere logo.
- Isolate the shaded portion of the page and rescan it with a different brightness setting to compensate for the shading.

These fixes require you to use *manual zoning.* You will recognize portions of the page, *specify zone contents* for the logo and text, and learn other manual zoning techniques in the course of this tutorial.

When to Use Manual Zoning

Use manual zoning in the following circumstances:

- to select a portion of a page for recognition
- to specify zone contents
- to order text for recognition
- to create a zone template for standardized pages

The next exercises cover the first three circumstances. Creating a zone template is covered in the next tutorial.

Manual Zones — Recognize Portions of a Page

You can recognize portions of a page to retain just the information you want to recognize and to leave out undesired elements.

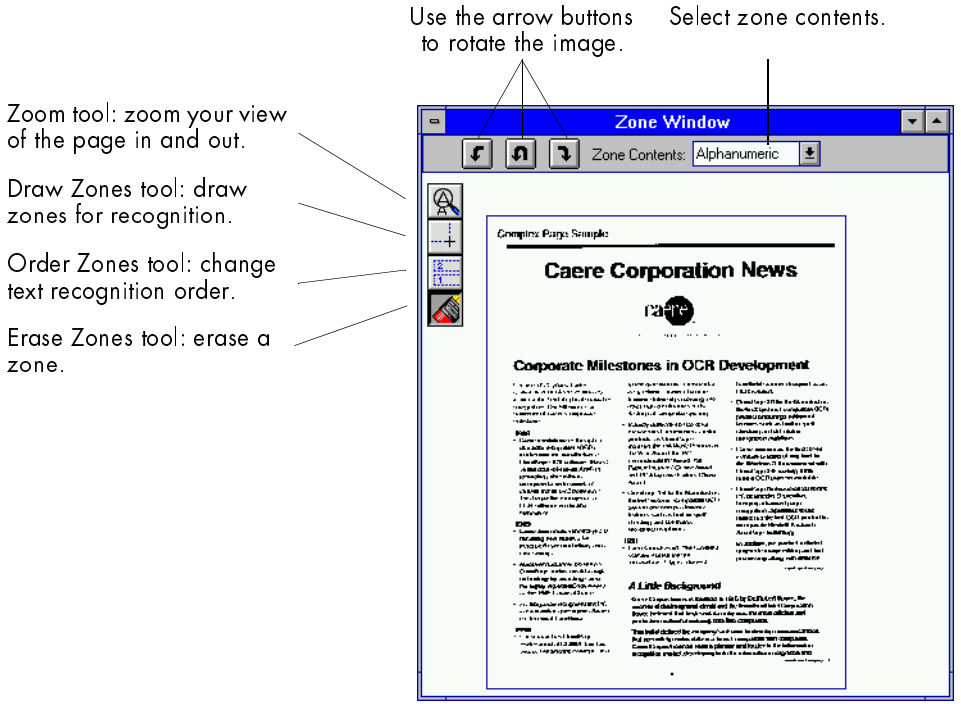
1. Choose *Tile Vertical* in the Window menu to view the text and zone windows side by side.

   You can also close the maximized text window and it will tile automatically with the zone window.

2. Select *Manual Zones* in the drop-down list under the Zone button in the toolbar.

3. Click *Yes* in the dialog box that asks if you want to replace the current zones.

The zones disappear and the automatic zone tools change to manual zone tools.



4 Click the Zoom tool.
Your cursor turns into a magnifying glass.

5 Click anywhere on the zone window to zoom into the image.
This is useful when you are drawing zones around areas that are close together such as the three columns on the page.

6 Click the right mouse button to zoom out of the image.

7 Click the Draw Zones tool.

8 Place the cursor by the Caere logo, hold down the mouse button, and drag the cursor to draw a rectangular zone around the title.
Leave out the volume number and other text below the logo.
OmniPage numbers this zone with a `1`.

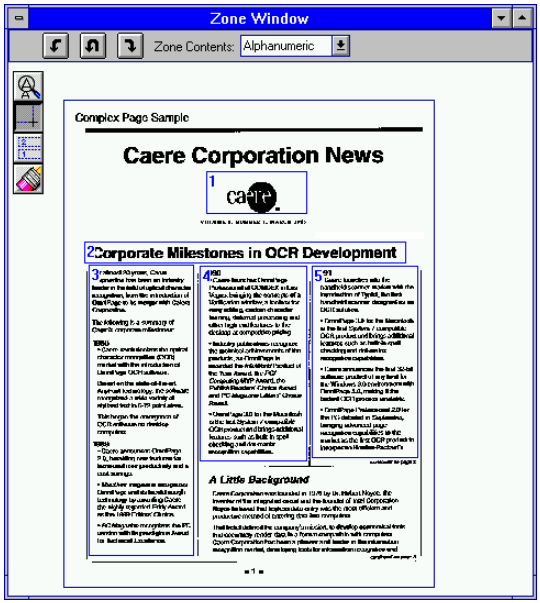
9 Draw a zone around the headline below the Caere logo.
OmniPage numbers this zone with a `2`.

10 Draw zones around the three side-by-side columns, avoiding the lines, as illustrated in the picture.

This is where zooming in your view of the page is especially of help.

Do not draw a zone around the *A Little Background* article. You will zone this separately in another exercise.

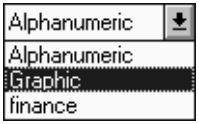
You should now have five zones as pictured below.



Manual Zones — Specify Zone Contents

1 Click in the zone around the Caere logo to make it active.

Handles appear on the zone when it is active.



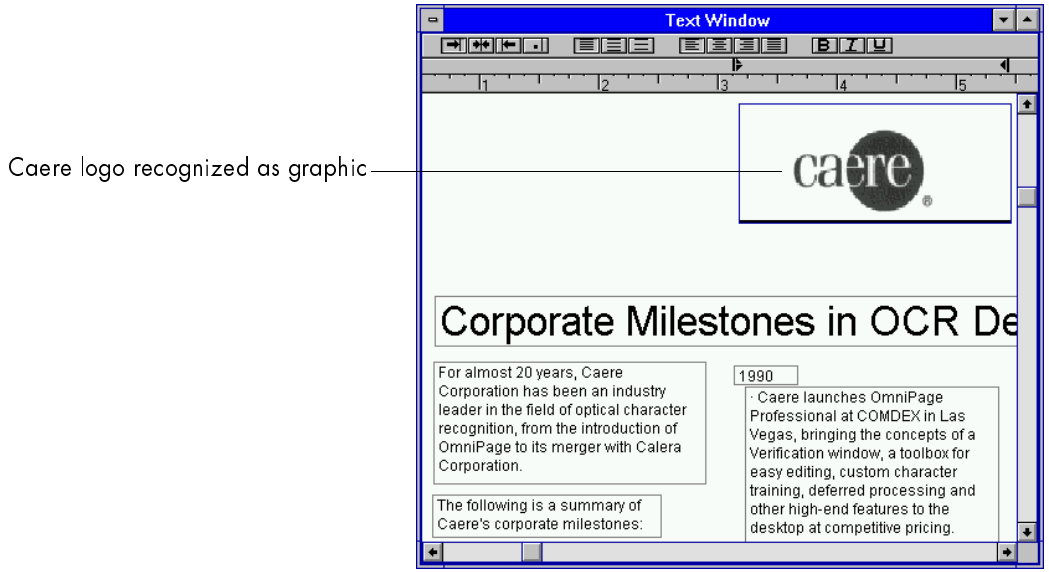
2 Select *Graphic* in the *Zone Contents* drop-down list.

The zone is now identified as a graphics zone. OmniPage will not try to recognize it as text.

3 Click the OCR button.

4 Click *Yes* in the dialog box that asks if you want to replace the current text.

OmniPage re-recognizes the document according to the zones you drew. The logo now appears in the text window as a graphic.



5 Leave the document open for the next exercise.

Manual Zoning — Reorder Text

After you scan a document, you may decide to reorder the text before or after recognition to save yourself time editing the document. In this exercise, you will recognize just the columns in a different order.

Reorder the Zones

1 Click the Erase Zones tool.

2 Click the first two zones to erase them.
OmniPage will recognize just the three columns.

3 Click the Order Zones tool.
The cursor becomes the # symbol and numbers in the three remaining zones around the columns disappear.

4 Click the right column that used to be labeled 5.
Now the zone is labeled 1. This zone will be recognized first and placed at the beginning of the new document in the text window.

5 Click the middle column.
It is now labeled 2 and will be recognized second.

6 Click the left column.
It is now labeled 3 and will be recognized third.

OCR

1 Click with your *right* mouse button on the OCR button to open the Settings Panel to the OCR options.

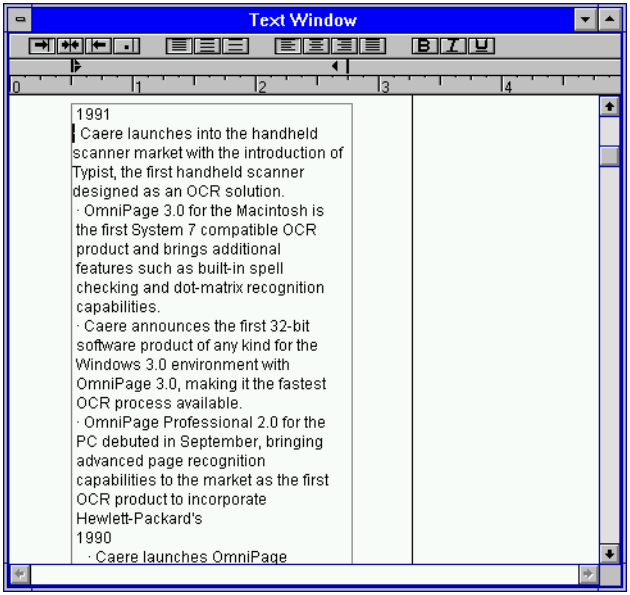
2 Select *Retain Font and Paragraph Formatting.*
This setting allows you to see the reordered text in the text window. Text would still be reordered with the True Page setting but you would have to export the text first and view it in the target application.

3 Click *Close.*

4 Click the OCR button.

5 Click *Yes* in the dialog box that asks if you want to replace the text.
OmniPage makes three recognition passes over the zones.

The text window opens to display the newly reordered text.



Scanning and the Brightness Setting

The scanner brightness setting you choose in the Scanner settings panel can strongly affect page recognition. *3D OCR with HP AccuPage 2/AnyPage* and *Auto Brightness with AnyPage/HP AccuPage 2* are both good scanner settings to choose for shaded areas.

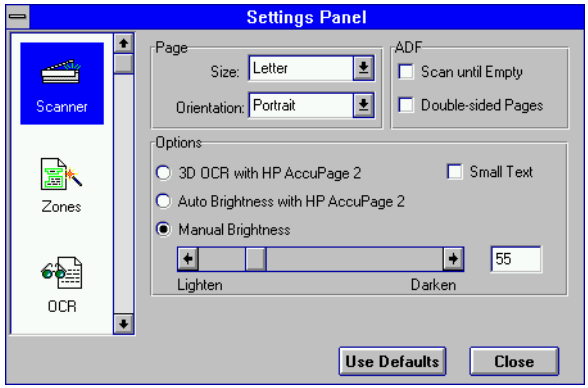
However, the shaded area in this case is too dark for the auto brightness settings to help much. You may find adjusting brightness manually works better. You will re-recognize just the *A Little Background* article to improve recognition and evaluate the article during processing.

Some scanners cannot scan a dark background well even with manual brightness adjustment. Skip this exercise if recognition does not improve after you have tried one or two different brightness settings.

1. Make sure the Complex Page sample is still in your scanner.
2. Select *Manual Zones* in the Zone button drop-down list if it is not already selected.

3 Click with your *right* mouse button on the Image button to open the Settings Panel to the Scanner options.

4 Select *Manual Brightness.*



The number range that appears in the text box on the right depends on what kind of scanner you have.

5 Drag the slider box to the left on the slider (toward *Lighten*).

You may have to experiment to find the optimum scanning brightness. For now, try to position the slider box approximately where it appears on the slider in the previous picture.

6 Note the number in the text box for future reference.

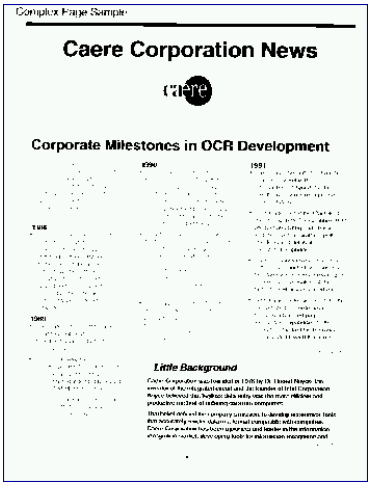
7 Click *Close.*

8 Click the Image button.

OmniPage rescans the document. It opens in the zone window as page two of your current document.

9 Look at the image to see how the brightness setting affected scanning.

Brightness setting too dark



Brightness setting too light

Brightness setting just right

- Set the brightness to a lighter setting if your image still has shading behind the article as does the left image, above, and rescan.
- Set the brightness to a darker setting if your image looks faded as does the middle image, above, and rescan.
- The right image, above, has the right brightness setting.

  The text outside the *A Little Background* article is lighter than the text inside the article (you can use the Zoom tool to enlarge the image and see).

  This would cause recognition problems if you recognized the entire page. That is why you will zone and recognize just the *A Little Background* article.

10 Choose *Delete Page* in the Edit menu to delete the page being viewed if it did not scan well.

Zone and Recognize the Article

1 Click the Draw Zones tool.

2 Draw a zone around just the *A Little Background* article.

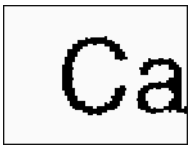
3 Click the OCR button.

4 Observe the Character window during OCR.

Shaded background dots would hinder recognition

Brightness setting too dark

Brightness setting too light

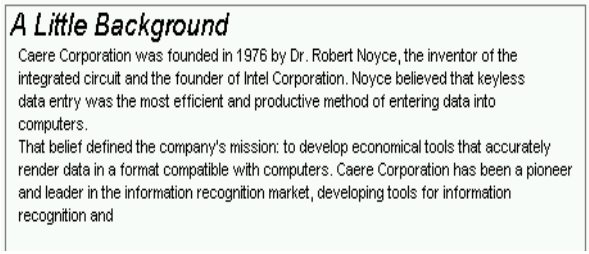


Brightness setting just right

- The Character window on the left, above, still shows some of the shaded background. Set the brightness to a lighter setting and rescan after OCR.
- The Character window in the middle, above, shows thin, broken characters. Set the brightness to a darker setting and rescan after OCR.
- The Character window on the right, above, shows well formed characters.

OmniPage displays the text in the text window after OCR.

5 Scroll down the page to locate the article in the text window.



A Little Background

Caere Corporation was founded in 1976 by Dr. Robert Noyce, the inventor of the integrated circuit and the founder of Intel Corporation. Noyce believed that keyless data entry was the most efficient and productive method of entering data into computers.

That belief defined the company's mission: to develop economical tools that accurately render data in a format compatible with computers. Caere Corporation has been a pioneer and leader in the information recognition market, developing tools for information recognition and

You should find few, if any, recognition errors once you have scanned with the proper brightness setting. Continue to adjust the scanner brightness setting in the Settings Panel and rescan the page if there are numerous errors.

6 Choose *Delete Page* in the Edit menu to delete the page being viewed if it did not scan or recognize well.

Cut and Paste the Text

If you deleted any pages that did not scan or recognize well, you should now have two pages with portions of recognized text on each. You can cut and paste the text in the *A Little Background Article* into the text from the rest of the newsletter.

1 Select the text in the *A Little Background* article.

2 Choose *Copy* in the Edit menu.

3 Click the left arrow button by the page number at the bottom of the window to go to page one.

This is the page that has the newsletter text from the three columns.

4 Place your cursor at the end of the text in the column.

5 Choose *Paste* in the Edit menu.

The text is added to the text on the page.

6 Resize the column as necessary to view all the text.

See “Working With Frames” on page 35 for detailed information on resizing and moving frames.

You could also export the whole document and cut and paste the text in your target application instead.

Tutorial 5 — Scanning a Single Column or Table

So far in these tutorials, you have scanned two different multiple-column documents with various settings. You may need to scan spreadsheets, tables, or memos. Although these also have multiple columns, these documents usually rely on tabs to maintain formatting. The *Single Column or Table* zoning method is specifically designed to recognize this sort of document.

You will use the *Single Column or Table* zoning method in this tutorial. You will also learn how to speed processing and increase recognition accuracy by creating a zone contents file and a zone template.

There are three exercises:

- Recognize a Memo With a Table
- Create a Zone Contents File
- Create a Zone Template

You will use the Single Column or Table Page sample in this exercise.

Recognize a Memo With a Table

1 Place the Single Column Page sample in your scanner making sure it is aligned correctly.

2 Click the drop-down lists under the process buttons and select:

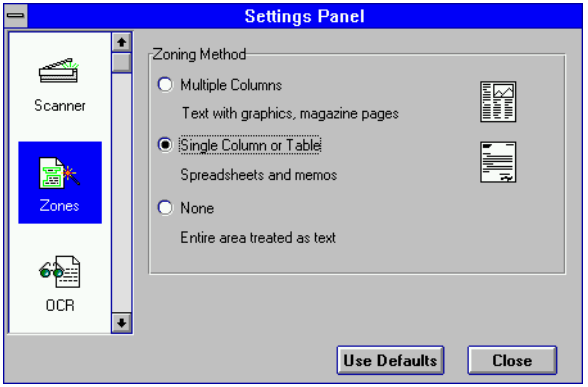
- *Scan Image*
- *Auto Zones*
- *Perform OCR*

3 Click the Settings Panel button in the toolbar.
The Settings Panel appears.

4 Click the *Zones* icon in the Settings Panel.

5 Select *Single Table or Column.*

This option is best for preserving tabbing or columns of characters such as are on the table on the sample page.

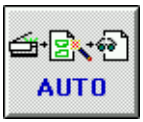


6 Click the OCR icon.

7 Select *Retain Font and Paragraph Formatting*.

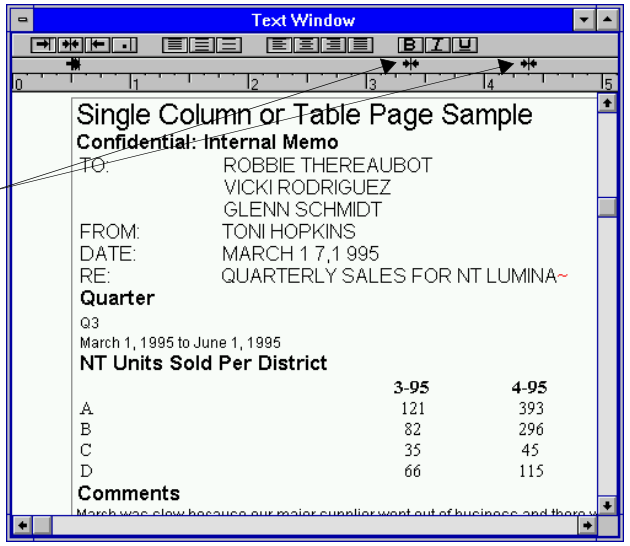
This option preserves the formatting of the page but not its layout as True Page would. On the sample page, for example, True Page would interpret the wide spacing between sections as extra line returns. You may not want this extra formatting.

8 Click *Close*.



9 Click *AUTO*.

OmniPage scans, zones, and recognizes the document.



Tabs inserted by OmniPage to maintain formatting

Note that OmniPage preserves the table and other even spacing with tabs.

The red tildes on the page mean OmniPage did not recognize some of the specialized characters in the document. You can double-click each tilde in the text window to open the Verification window and see the original image. A later tutorial, Train OCR, shows you how to teach OmniPage to recognize these special characters and symbols.

10 Leave the document open for the next exercise.

Create a Zone Contents File

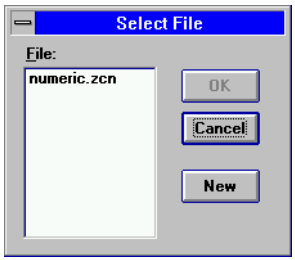
You can speed OCR and minimize potential recognition errors by creating your own zone contents file. Depending on the font and image quality, OmniPage may recognize a five (5) as an *S* or a zero (0) as an *O*. A zone contents file prevents this by telling OmniPage exactly what to recognize in a particular zone. You will create a zone contents file in this exercise.

Creating the File

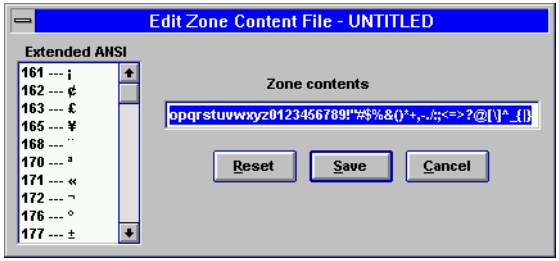
1 Choose *Edit Zone Contents File...* in the Settings menu.

The Select File dialog box opens.

2 Click *New*.



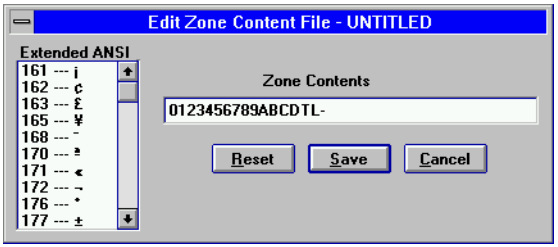
The Edit Zone Content File dialog box opens with a string of highlighted characters. This is the default Alphanumeric zone contents set



You need to enter all the numbers in the table. You must also enter any characters. If you just entered numbers, OmniPage would not be able to recognize the letters with this zone contents file.

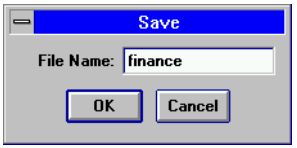
3 Type the characters `0123456789ABCDTL-` (hyphen).

Zone contents files are case-sensitive, so make sure your letters are uppercase as in the example.



The highlighted characters are replaced with the ones you enter.

4 Click *Save*.
The Save dialog box opens.



5 Type `finance` in the *File Name* text box.

6 Click *OK*.

Draw and Specify Zones

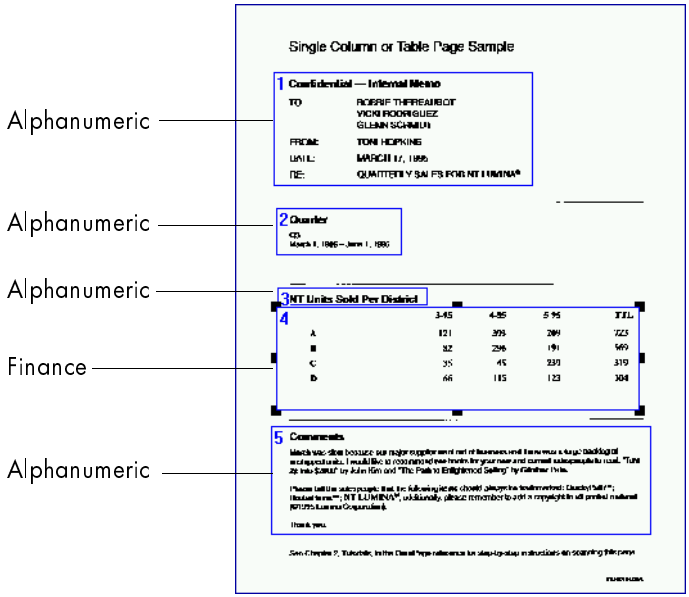
1 Follow steps 1–8 beginning on page 58 if you did not leave the document open.

2 Choose *Tile Vertical* in the Window menu so you can see the zone window.

3 Select *Manual Zones* in the Zone button drop-down list.

4 Click *Yes* in the dialog box that asks if you want to delete the current zones.

5 Draw zones around the sections of the page as shown in this picture:



6 Click in the zone around the table to make it active.

7 Select *Finance* in the *Zone Contents* drop-down list.

8 Click the OCR button.

9 Click *Yes* in the dialog box that asks if you want to replace the current text.

OmniPage recognizes each of the zones according to the zone contents you specified.

Because you selected the appropriate zone contents file, all characters in the table are recognized correctly.

10 Leave the document open for the next exercise.

Create a Zone Template

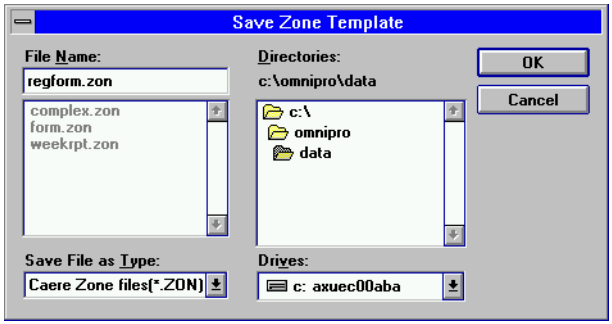
The Single Column or Table Page sample is a fictional example of a weekly report — one that always has similar information in the same place on the page. This is known as a *standardized form*. You can create a zone template to use on standardized form instead of drawing the same zones each time.

Select Settings

1 Perform the previous exercise, “Draw and Specify Zones” if you did not leave the document open.

2 Choose *Save Zone Template...* in the File menu.

The Save Zone Template File dialog box appears.



3 *Caere Zone (`*.zon`)* is the only selection in the *Save Files as Type* drop-down list.

4 Type the name `weekrpt.zon` in the *File Name* text box.

The `data` directory is the default location for all zone template files. This is where OmniPage looks for them. It cannot find the zone template files in any other location.

5 Click *OK*.

Load the Zone Template

1 Select *weekrpt* in the Zone button drop-down list.

2 Click *Yes* in the dialog box that asks if you want to replace the current zones.

3 Click the Zone button.

OmniPage draws zones on the page image according to the zone template you just saved.

4 Click each zone and observe the setting in the Zone Contents drop-down list to verify that your zone template is correct.

You could use this template on any similar documents.

You can create zone templates for any page that has a standardized layout. You could also load a saved settings file before OCR so that an OCR training file could be used on the document. See “Save a Settings File” on page 37.

The next tutorial, “Train OCR,” teaches you how to create a training file.

Tutorial 6 — Train OCR

OmniPage automatically recognizes characters commonly found in most documents. Other documents may contain characters OmniPage has not yet learned to recognize such as copyright and trademark symbols, and mathematical symbols such as pi (π). You can train OmniPage to recognize special characters and create a training file to use on similar documents.

This tutorial contains the following sections:

- Scan a Document With Special Characters
- Train OCR to Recognize Special Characters

You will use the Single Column or Table Page sample in this exercise.

Scan a Document With Special Characters

1 Place the Single Column or Table Page sample in your scanner making sure it is aligned correctly.

2 Click the drop-down lists under the process buttons and select:

- *Scan Image*
- *week.rpt*

  This template was created in the last tutorial. Select *Auto Zones* if you did not perform the last tutorial.
- *Perform OCR*

3 Click the Settings Panel button in the toolbar.

The Settings Panel appears.

4 Click the *Zones* icon in the Settings Panel.

5 Select *Single Table or Column*.

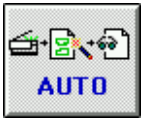
This option is best for preserving the tabbed spacing found on the sample page.

6 Click the OCR icon.

7 Select *Retain Font and Paragraph Formatting*.

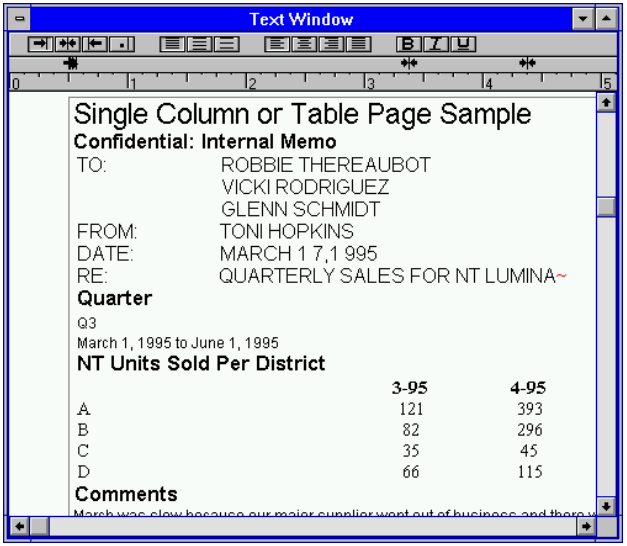
You do not need to retain exact page layout in this exercise.

8 Click *Close*.



9 Click *AUTO*.

OmniPage scans, zones, and recognizes the document. and then displays the recognized text in the text window.



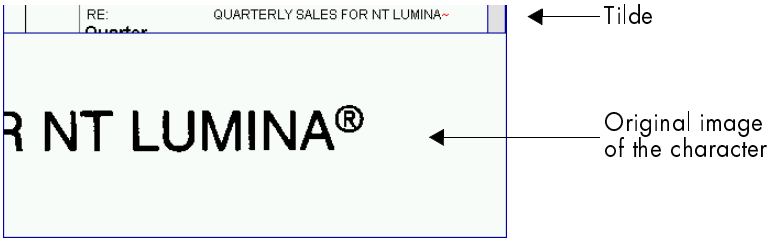
View the Recognized Text

1 Compare the text in the text window to the page you scanned.

OmniPage replaced unrecognizable characters with red tildes.

2 Double-click a red tilde if you have any, such as the one after the word LUMINA in the example above.

The Verification window opens to show the original scanned character, a registered trademark sign.



You will train OCR to recognize this and other characters.

3 Click anywhere outside the Verification window to close it.

Leave the document open. You will create a training file in the next exercise.

Train OCR to Recognize Special Characters

You will train OmniPage to recognize several characters in this exercise. See "Scan a Document With Special Characters" on page 65 if you did not leave the document open.

Re-recognize the Document

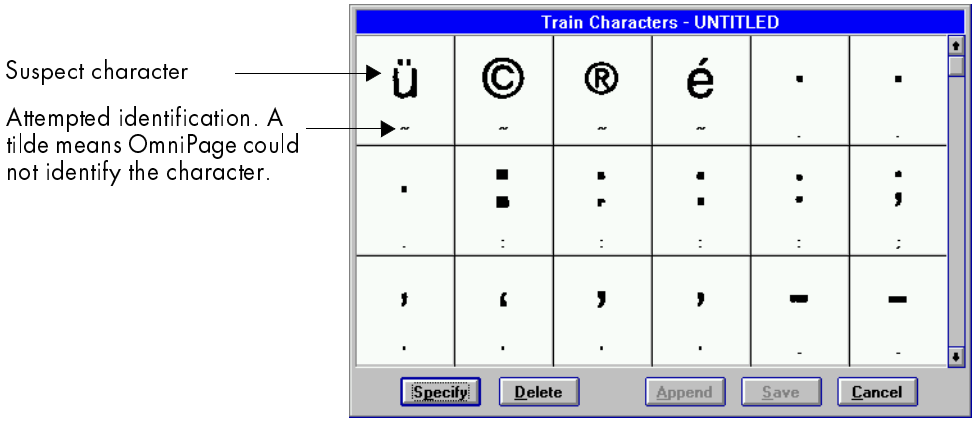
1 Select *Train OCR* in the OCR button drop-down list.

2 Click the OCR button.

3 OmniPage re-recognizes the document, and then opens the Train Characters dialog box.

The Train Characters Dialog Box

Characters OmniPage had trouble identifying are displayed at the top of the dialog box.



Beneath each image, in smaller type, is OmniPage's attempted identification of that character. A tilde means that OmniPage could not identify the character.

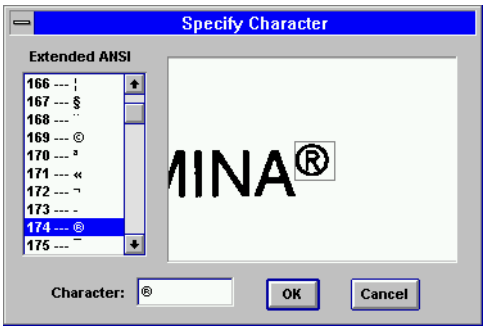
Depending on your scanner, the characters you see in this dialog box may be different from those pictured above.

Specify Characters to Recognize

1 Locate the registered trademark (®) symbol in the dialog box.

2 Double-click the symbol, or select it and click *Specify*.

The Specify Character dialog box appears.



It displays the symbol as it appeared in the scanned document.

3 Locate the registered trademark symbol in the *Extended ANSI* list box on the left.

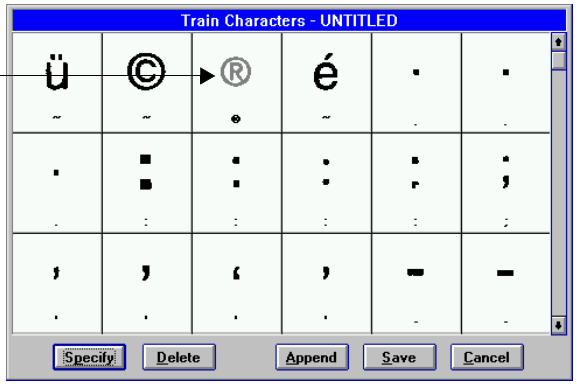
4 Double-click the symbol.
It appears in the *Character* edit box.

If a symbol or character does not appear in the list, you can type it in the *Character* edit box, cut and paste it from another source, or use an Alt-number key combination.

---

5 Click *OK*.
The specified character now appears under the suspect character in the Train Characters dialog box.

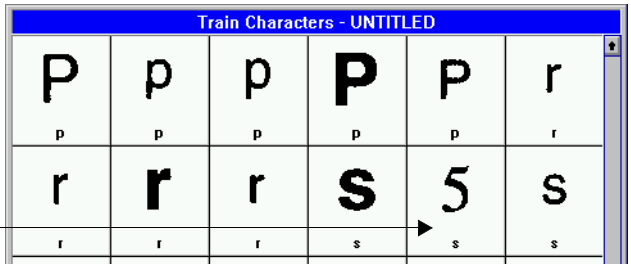
Tilde replaced with registered trademark symbol



The symbol turns gray to indicate that you specified a character for it.

6 Specify other characters in the same way such as the lowercase ü and é, and the copyright (©) symbol.

Characters OmniPage believes it identified correctly are listed alphabetically below the suspect characters. Check for common errors, such as a *5* being recognized as the letter *S*.



5 recognized as the letter *S*. You can create and use a zone contents file to prevent this (see "Create a Zone Contents File" on page 60 for information).

Generally, you will not want to train OmniPage to recognize common letters unless they are in a very specialized font. One way to prevent these common errors is to use a special zone contents files during recognition, such as the numeric *finance* zone contents file created in the previous tutorial.

Even if a character is not recognized, OmniPage corrects most common OCR errors by analyzing the structure of a word and comparing it to entries in the dictionary.

Save the File and Recognize the Document

1 Click the *Save* button.

The Save dialog box appears.

2 Type a file name in the *File Name* edit box.

3 Click *OK*.

A dialog box asks if you want to recognize the image with the training file you just created.

4 Click *Yes*.

OmniPage recognizes the document and all specified symbols.

5 Check the text window to verify the training file improved OCR.

6 Close the document and click *No* in the dialog box that asks if you want to save changes.

This file becomes the default training file in the OCR section of the Settings Panel. You can save Settings Panel selections, including this training file, as a settings file for use on similar documents. See "Save a Settings File" on page 37. See "Train OCR" on page 129 for more information on creating and editing an OCR training file.

Tutorial 7 — Deferring OCR

Compared to the time it takes to scan and zone a page, OCR can be time-consuming. You might find it more efficient to scan a stack of pages (especially if you have an ADF) or load multiple images all at once, zone them, and then defer recognition to a later time.

You can choose to finish OCR at any time convenient to you or set it to take place automatically at a specific time. In this tutorial, you will scan two pages, defer OCR, and then finish OCR both on an open document and on a saved document.

Use the Quick Scan Page sample and the True Page sample in this tutorial.

This tutorial contains the following exercise:

- Scan Multiple Pages and Defer OCR
- Finish Current Document
- Finish Deferred Documents

Scan Multiple Pages and Defer OCR

1 Click the drop-down list under each process button and select these options:

- *Scan Image*
- *Auto Zones*
- *Defer OCR*

2 Click the Settings Panel button or choose *Settings Panel...* in the Settings menu.

The Settings Panel appears.

3 Click *Use Defaults.*

4 Click *Yes* in the dialog box that asks if you are sure.

5 Make the following selection if you are using an automatic document feeder (ADF):

- Click the Scanner icon.
- Click *Scan until Empty*.

6 Click *Close.*

7 Place both sample pages in your ADF.

If you do not have an ADF, place the Quick Scan Page sample in your scanner.

8 Click the *AUTO* button.

The first page in the stack is scanned and zoned, and then the next page.

If you do not have an ADF, place the True Page sample in your scanner now and click *AUTO*.

You now have a two-page document open in the zone window.

9 Leave the document open for the next exercise.

You have two choices at this point: finish recognizing the current open document or save the document and perform recognition later. You will finish the current open document in the next exercise.

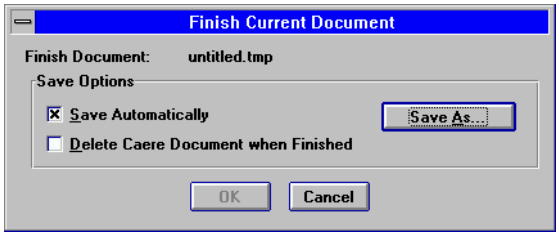
Finish Current Document

In the normal course of a day, you may decide to leave scanned or loaded documents open in OmniPage and finish them later. OCR can be both time- and memory-intensive.

You can even set OCR to begin and leave your computer while it is in process.

1 Choose *Finish Current Document...* in the Process menu.

The Finish Current Document dialog box appears.



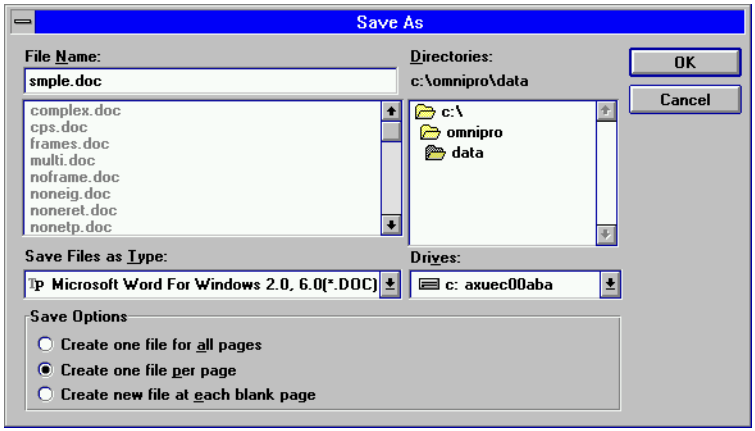
You can perform recognition and save the document later, or perform recognition and save the document automatically. You will save the document automatically in this exercise.

2 Select *Save Automatically* if it is not selected.

This activates the other options in the dialog box.

3 Click *Save As...*.

The Save As dialog box appears.



4 Select a file type in the *Save Files as Type* drop-down list.
Microsoft Word for Windows is selected in the example above.

5 Select a location for your saved file.

6 Select *Create one file per page.*
This save option creates two separate files after OCR. See "Save Options" on page 95 for information on the other two save options.

7 Type `smple` in the *File Name* text box.
You can type in a name of up to five characters, not including the extension, with this save option.

8 Click *OK* to return to the Finish Current Document dialog box.

9 Click *OK* to begin OCR.
OmniPage recognizes each page and saves it as specified. The Caere Document remains open in OmniPage with the recognized text displayed in the text window.

10 Choose *Close Document* in the File menu, saving changes to the Caere Document if you wish.

You now have two new files in the directory you selected. The Quick Scan sample is named `smple001.*`. The True Page sample is named `smple002.*`. OmniPage has appended the appropriate file extension if you did not type it in. (In this example, the full file names are `smple001.doc` and `smple002.doc`.)

See "Finish Current Document" on page 133 for detailed information on this command.

Finish Deferred Documents

You may decide to defer OCR, close the open documents or OmniPage, and finish processing later. You must save the open documents in order to reopen and recognize them.

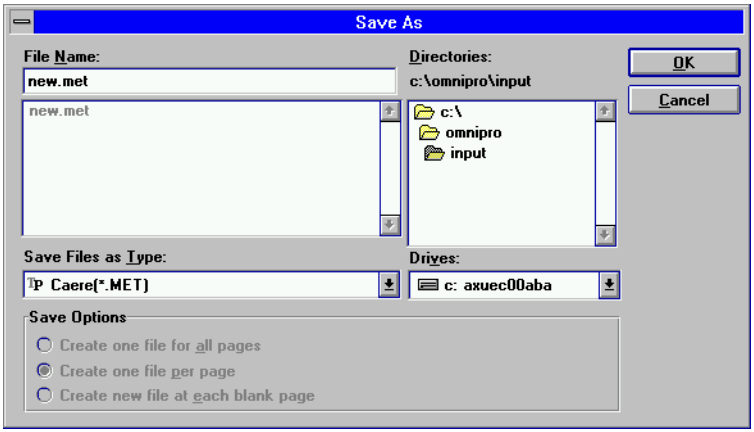
Scan and Save the Pages

1 Follow the steps in the section "Scan Multiple Pages and Defer OCR" on page 70 and then return to this section.

OmniPage scans and zones the pages. You now have a two-page document open in the zone window.

2 Choose *Save As...* in the File menu.

The Save As dialog box appears.



3 Locate the `omnipro\input` directory.

This is the default location in which OmniPage looks for deferred files.

4 Select *Caere[*.MET]* in the *Save Files as Type* drop-down list.

You can only select Caere Documents and image files when finishing deferred OCR.

5 Type `new.met` in the *File Name* text box.

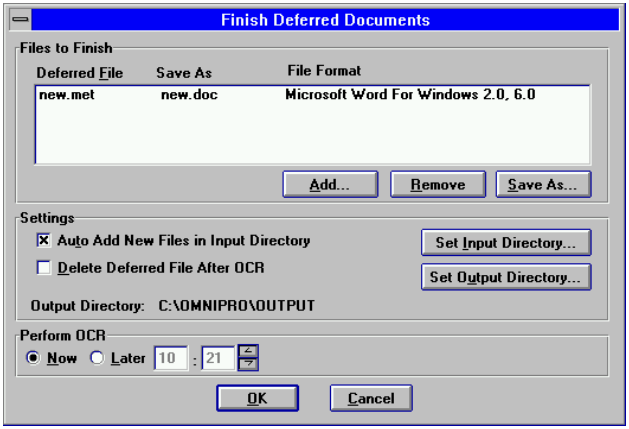
6 Click *OK*.

7 Choose *Close Document* in the File menu to close the zone window.

The Finish Deferred Documents Dialog Box

1 Choose *Finish Deferred Documents...* in the Process menu.

The Finish Deferred Documents dialog box appears.



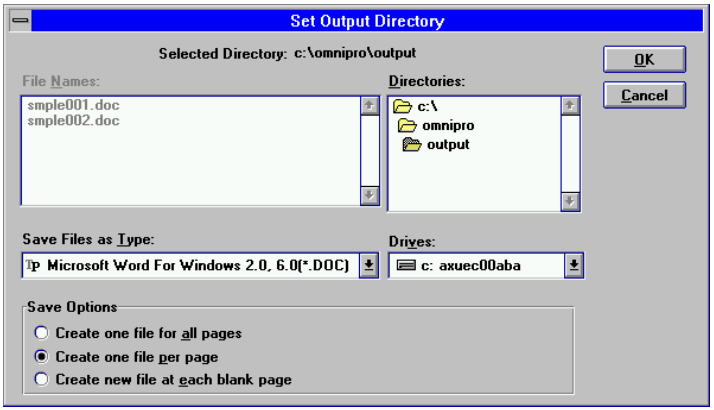
2 The file you saved to the `input` directory appears in the *Files to Finish* list box.

This is where OmniPage looks by default for deferred files. OmniPage assigns a a file format to your file based on the last-selected file format. (In the previous example, Word for Windows was the selected file format so it is selected in this example by default.) It assigns a new file name based on that file format.

You can select a different location and file format for files if you wish.

3 Click *Set Output Directory...*.

The Set Output Directory dialog box appears.



(A *Network* button also appears in this dialog box if you use Windows For Workgroup with network enabled.)

- Locate and select the `output` directory as the location to save the file if it is not already selected by default.
- Select *Create one file per page* if it is not already selected.
- Select a file format in the *Save Files as Type* drop-down list if you want to change the current selection.

In this exercise, there is just one file selected for OCR. Note that if you select multiple deferred documents, however, selections made in the Set Output Directory dialog box would affect *all* files.

4 Click *OK* to return to the Finish Deferred Documents dialog box.

Perform OCR

1 Deselect *Delete Deferred File After OCR* under *Settings* if it is selected.

2 Select *Now* under *Perform OCR*.

3 Click *OK*.

OmniPage opens and recognizes the `new.met` file, and then saves it as specified. You now have three more new files in the `output` directory.

The Quick Scan sample is named `new001.*`. The True Page sample is named `new002.*`. OmniPage has appended the appropriate file extension if you did not type it in. (In this example, the full file names would be `new001.doc` and `new002.doc`.)

The `new.met` file also moves here after OCR. If you had selected *Delete Deferred File After OCR*, OmniPage would have deleted the `sample.met` file permanently from the `input` directory instead.

See "Finish Deferred Documents" on page 135 for detailed information on this command.

Tutorial 8 — Using Direct Input

You will use OmniPage's Direct Input mode in this tutorial to scan and recognize text from within another application. Recognized text will be pasted directly into the initiating application.

This tutorial consists of one exercise containing the following sections:

- Register an Application
- Launch Direct Input
- Direct Input Mode

You will use the Quick Scan Page sample in this tutorial.

Register an Application

You must *register* an application before using it to initiate Direct Input. Once an application is registered with Direct Input, the *Direct Input...* command appears in its File menu above the *Exit* command. You choose this command to initiate OCR processing from your application.

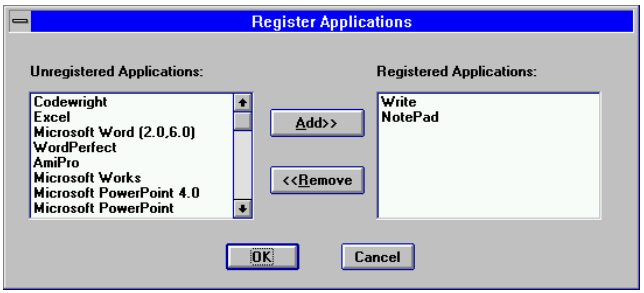
A variety of applications are compatible with Direct Input. You will register a compatible application in this exercise if you have one.

1 Launch OmniPage if it is not already open.

2 Choose *Register Applications...* in the Settings menu.

This command is enabled only when *Enable Direct Input* is selected in the Direct Input settings panel.

The Register Applications dialog box appears.

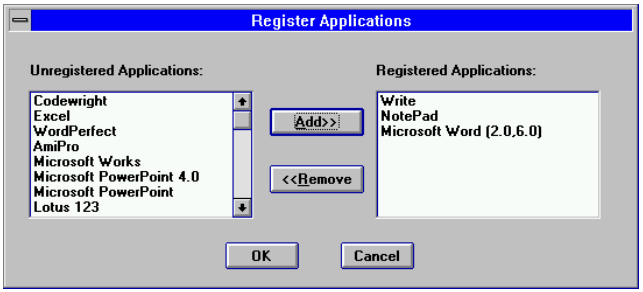


The Windows programs Write and NotePad are pre-registered.

3 Select an application that is installed on your computer in the *Unregistered Applications* list box.

4 Click *Add>>*.

The application moves into the *Registered Applications* list box.



5 Select and move as many applications as you like.

6 Click *OK* when you are done.

OmniPage immediately places the *Direct Input...* command in the File menu of the registered application(s).

7 Choose *Exit* in the File menu.

Launch Direct Input

1 Place the Quick Scan Page sample in your scanner making sure it is aligned properly.

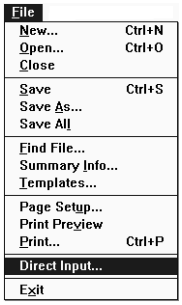
2 Open or switch to any registered application.

Microsoft Word is used in this example.

3 Use your program's commands to create a new document if one is not open.

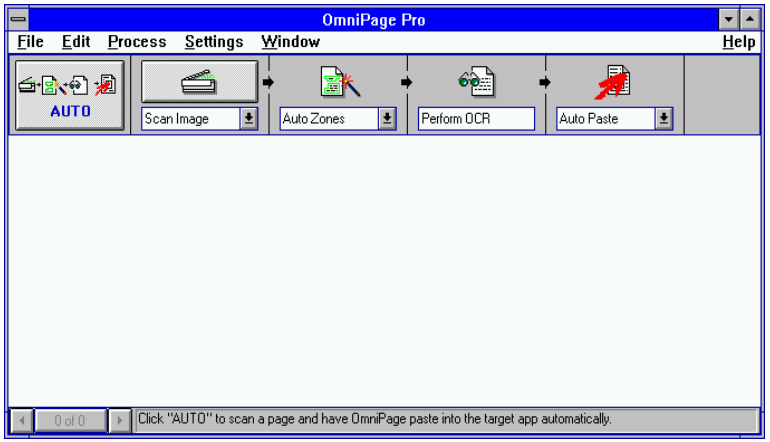
4 Place your cursor in this new document if it is not already there.

5 Choose *Direct Input...* in the program's File menu.



Some applications, such as Word and Notepad, allow you to launch multiple copies of the application. The *Direct Input...* command only appears in the first copy of the application launched.

OmniPage launches in Direct Input mode and the Direct Input window appears.



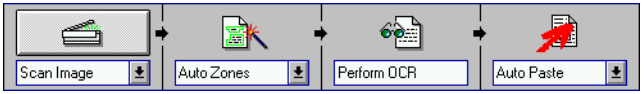
Direct Input Mode

Always select the appropriate settings before you begin the OCR process.

1 Click the drop-down list under each process button and select these options:

- *Scan Image*
- *Auto Zones*

- *Auto Paste*

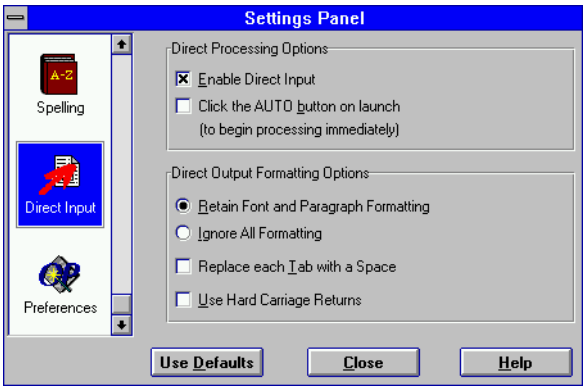


Perform OCR is the only selection under the OCR button.

2 Choose *Settings Panel...* in the Settings menu to open the Settings Panel.

There are no shortcut command buttons in Direct Input mode as there are in the regular OmniPage mode.

3 Click the Direct Input icon to observe the settings.

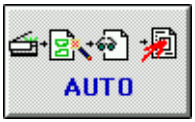


4 Click *Use Defaults.*

5 Click *Yes* if a dialog box appears to confirm your choice.

The default output formatting option is *Retain Font and Paragraph Formatting.* This setting retains font types and styles, and paragraph order and formatting in recognized text.

6 Click *Close.*



7 Click *AUTO.*

You can click *STOP* at any time to cancel processing but remain in Direct Input mode.

OmniPage scans, zones, and recognizes the document in the Direct Input window. Then the program exits, your initiating application appears, and the text is pasted where you left the cursor.

If you had for some reason closed your initiating application or had not opened or created a document, OmniPage would paste the recognized text to the Clipboard instead. Use your program's commands to paste text from the Clipboard into the application of your choice.

See Chapter 5, Direct Input, for detailed information on this feature.

Chapter 3

Commands and Settings

This chapter explains how to use the OmniPage commands and settings, all of which are located within eight menus and a toolbar.

This chapter contains the following sections:

- The Toolbar
- The File Menu
- The Edit Menu
- The Format Menu
- The Process Menu
- The Settings Menu
- The Register Menu*
- The Window Menu
- The Help Menu

The menu command information is listed in the same order in this chapter that the commands appear in the menu.

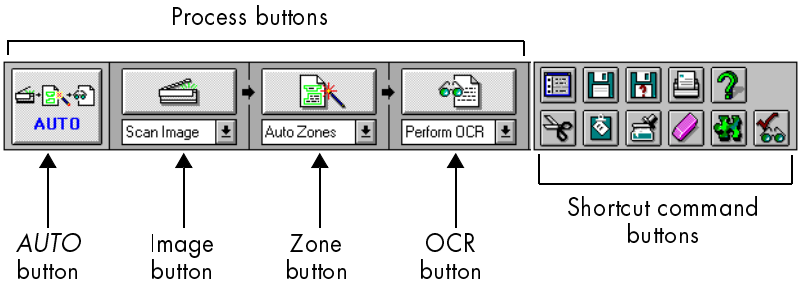
See Chapter 5, Direct Input, for an explanation of the different toolbar options and menu commands available in Direct Input mode.

Many of the operations explained in this chapter are detailed further in the form of tutorial exercises. Please refer to Chapter 2, Tutorials, for information on basic and advanced document processing.

* The Register menu only appears if you did not register your copy of OmniPage the first time you launched it after installation.

The Toolbar

The toolbar has four process buttons and several shortcut command buttons.



Use the toolbar to access the three basic steps of the optical character recognition (OCR) process:

1 Acquiring a page image to recognize.

2 Creating zones on the image to choose what will be recognized.

3 Performing OCR on the information in the zones.

OCR is the process of converting an image file to editable text. An image is an electronic picture of text and / or graphics. You acquire an image by scanning a hard-copy document or loading a graphic-format file (such as a TIFF or PCX file). The image you scan or load is just a picture to your computer before OCR.

During OCR, OmniPage looks for and defines characters on the image to produce editable text. You can export the recognized text from OmniPage for use in a wide variety of applications.

The toolbar's process buttons perform the same operations as the *Process Settings* commands in the Process menu. The shortcut command buttons provide shortcuts for performing other OmniPage commands.

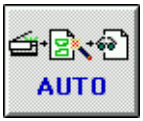
Click the:

- *AUTO* button to process your document automatically from start to finish according to the selected processing commands.
- Image button to acquire an image for recognition by scanning a page or loading an existing image.
- Zone button to specify what will be recognized in an image by creating zones manually, automatically, or with a template.
- OCR button to perform OCR, defer OCR, or train OCR.
- Shortcut command buttons to access various menu commands.

The toolbar is different in Direct Input mode. See Chapter 5, Direct Input, for detailed information.

Each button is described next in the order it appears on the toolbar.

AUTO Button



The *AUTO* button is the first button in the toolbar. It performs the same operations as the *Auto* command in the Process menu.

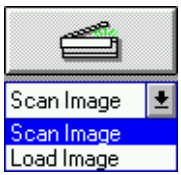
Click *AUTO* to start and finish processing each page of a new document or to finish processing the current page of an open document according to the currently selected Process Settings commands. This is known as *automatic processing*.

For example, if you select *Scan Image*, *Auto Zones*, and *Perform OCR* in the processing button drop-down lists and click *AUTO*, the first page in the scanner is scanned, automatically zoned, and recognized. You do not have to click each process button individually.

You can also click *AUTO* to finish processing the current page of an open document. The resulting operation depends on the state of the page and the selected Image, Zone, and OCR commands. If the page image is zoned and you click *AUTO*, for example, then OmniPage immediately begins OCR according to the selected OCR button command.

The *AUTO* button changes to *STOP* when automatic processing begins. Click *STOP* at any time if you want to discontinue processing.

Image Button



The Image button is the second button in the toolbar. This button contains the same commands, *Scan Image* and *Load Image*, that are in the cascading menu under the *Process Settings* command in the Process menu.

Click the Image button to acquire an image by scanning a page or loading an existing image file. The two commands are described further in this section.

OmniPage uses the selected Image button command when it performs automatic processing.

Use your right mouse button to click the Image button and automatically open the Settings Panel to Scanner options.

Scan Image

Select *Scan Image* to scan a page in your scanner. This command only appears in the drop-down list if you have installed the Scan Manager. Select your default scanner in the Scan Manager before scanning (see “Scan Manager Installation” on page 8). Select the appropriate Scanner options in the Settings Panel as well.

A progress meter appears and the status bar reports progress during scanning. The page image appears in the zone window when scanning is complete.

Click the *STOP* button in the toolbar to cancel scanning at any time.

Load Image

Select *Load Image* to load a previously saved image file as a new document or to add it as a new page to an open document.

An image file is a picture of text and / or graphics that is saved in an image file format such as TIFF or PCX. When you load an image file in OmniPage, it appears in the zone window. See “Supported Input File Formats” on page 239 for a list of files OmniPage can load.

Click the *STOP* button in the toolbar to cancel processing at any time. See “Load Image” on page 119 for detailed information on this command.

Zone Button

The Zone button is the third button in the toolbar. This button contains the two commands, *Auto Zones* and *Manual Zones*, that are in the cascading menu under the *Process Settings* command in the Process menu. (The Zone button drop-down list contains the names of available zone templates rather than the Process menu command *Use Template...*.)

Click the Zone button to create zones that determine what will be recognized in the page image. The available commands are described further in this section.

OmniPage uses the selected Zone button command when it performs automatic processing.

Use your right mouse button to click the Zone button when it is active and automatically open the Settings Panel to Zones options.

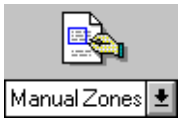
Auto Zones

Select *Auto Zones* in the drop-down list to have OmniPage automatically draw and order zones for text recognition on the current page image.

OmniPage uses the selected Zones option in the Settings Panel: *Multiple Columns*, *Single Column or Table*, or *One Zone*. For more information about each of these options, see "Zones Options" on page 163.

If the current page already has zones when you select this command, you are prompted to delete the current zones before auto zoning occurs. Click *Yes* to have OmniPage delete old zones and draw new zones.

Manual Zones



Select *Manual Zones* to draw and order your own zones for text recognition on the current page image.

OmniPage uses the selected Zones option in the Settings Panel on the zones you draw: *Multiple Columns*, *Single Column or Table*, or *One Zone*. For more information about each of these options, see "Zones Options" on page 163.

If the current page already has zones when you select this command, you are prompted to delete the current zones. Click *Yes* to have OmniPage delete old zones so that you can draw new zones.

For more information on creating manual zones, see "Tutorial 4 — Evaluating a Page" on page 44.

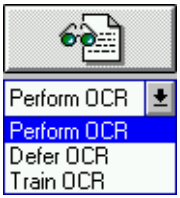
Zone Templates

Select a previously created zone template file to automatically zone the current page image. Zone template files appear in the Zone button drop-down list after they are saved. A template contains zones that you created manually for a page and then saved as a file along with the zones' order, position, and contents.

Using a zone template is a quick and efficient means of processing documents that have the same zoning requirements. See "Save Zone Template" on page 100 for detailed information on creating zone templates.

If the current page already has zones when you select a template, you are prompted to delete the current zones. Click *Yes* to have OmniPage delete old zones and apply the selected zone template.

OCR Button



The OCR button is the fourth button in the toolbar. This button contains the same commands, *Perform OCR*, *Defer OCR*, and *Train OCR*, that are in the cascading menu under the *Process Settings* command in the Process menu.

Click the OCR button to perform the selected OCR command on the page image. The available commands are described further in this section.

OmniPage uses the selected OCR button command when it performs automatic processing. Zones are created automatically if you click the OCR button before clicking the Zone button or before drawing manual zones.

Use your right mouse button to click the OCR button when it is active and automatically open the Settings Panel to OCR options.

Perform OCR

Select *Perform OCR* to recognize text on the current page.

Before performing OCR, make sure the appropriate OCR options are selected in the Settings Panel.

If there are no zones on the page when you select *Perform OCR* and click the OCR button, OmniPage automatically creates zones according to the selected Zone command. If *Manual Zones* is currently selected, OmniPage ignores this and draws zones automatically.

Defer OCR

Select *Defer OCR* to delay text recognition of one or more pages of your document. OCR can be a time- and memory-intensive process so you may want it to take place while you are away from your computer.

You might, for example, choose *Scan Page*, *Auto Zones*, and *Defer OCR* as the processing commands. OmniPage will scan and zone the document and stop processing it further. Save deferred documents as Caere Documents (`*.met`).

Choose *Finish Current Document* or *Finish Deferred Documents* in the Process menu when you want to perform page recognition on the deferred document(s).

See “Finish Current Document” on page 133 and “Finish Deferred Documents” on page 135 for detailed information.

Train OCR

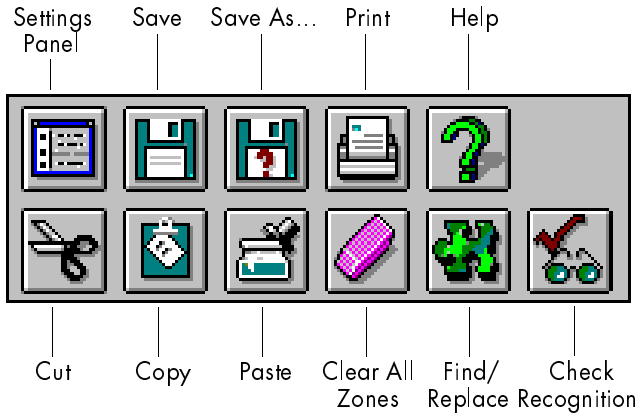
Select *Train OCR* to create a character training file (`*.trn`) that assists OmniPage during text recognition and allows better recognition of special characters.

A character training file is a set of pre-recognized text characters that OmniPage compares with the characters in the page image during recognition. Before recognizing an image, you can create a new training file or choose an existing one in the OCR settings panel.

For more information on creating a training file, see “Train OCR” on page 129.

Shortcut Command Buttons

The shortcut command buttons perform the same functions as the commands of the same name in the File, Edit, Settings, and Help menus.

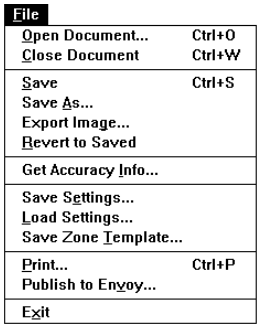


For example, you can click the Settings Panel button in the toolbar to open the Settings Panel or you can choose *Settings Panel...* in the Settings menu.

See each button’s respective menu entry further in this chapter for information.

Some buttons are only active when the particular command it represents can be applied to the active text or zone window. The Check Recognition button, for example, is only active when the text window is active. There are no Shortcut command buttons in Direct Input mode. See Chapter 5, Direct Input, for detailed information.

The File Menu



The File menu lets you manage OmniPage file operations. File menu commands include:

- Open Document
- Close Document
- Mail (MAPI mail systems only)
- Save
- Save As
- Export Image
- Revert to Saved
- Get Accuracy Info
- Save Settings
- Load Settings
- Save Zone Template
- Print
- Publish to Envoy
- Exit

Open Document

Choose *Open Document...* to open a Caere Document (`*.met`) or an image file. A Caere Document is created the first time you scan a page or load an image file. This is a proprietary OmniPage file format. See “Caere Document (*.met)” on page 92 for more information.

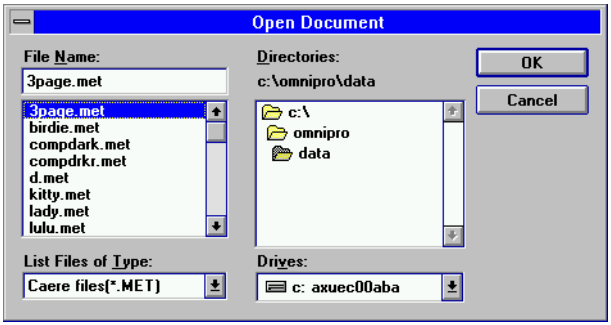
Image file

An image file is a “picture” of text and/or graphics that is saved in an image file format such as TIFF or PCX. Received fax files, for example, can usually be saved in an image format OmniPage recognizes. Image files do not have OCR or zone information. When you open an image file in OmniPage, it appears in the zone window.

Opening a Caere Document or Image File

1 Choose *Open Document...* in the File menu.

The Open Document dialog box appears.



2 Select the type of file to open in the *List Files of Type* drop-down list.

Files of that type appear in the *File Name* list box.

3 Double-click a file or select it and click *OK*.

The image file opens in the zone window. A Caere Document opens with recognized text in the text window (if it was recognized) and its original image in the zone window. In either case, the first page of your file is displayed.

Click *Cancel* to exit without opening a file.

An image file becomes a Caere Document once it is opened with the *Open...* command. You can only open one Caere Document at a time. OmniPage closes the current document if you open another one. It prompts you to save the current document if you have made changes to it.

Add page images to your open document by choosing *Load Image* or *Scan Image* in the Process menu or in the Image button drop-down list. See “Adding a Page to a Scanned Image” on page 119 and “Adding a Page to a Loaded Image” on page 121 for detailed information.

Close Document

Choose *Close Document* to stop working on a document but leave OmniPage running.

If the current document has not been saved or has changed since the last save, a prompt appears asking if you want to save the document before closing. See “Save As” on page 90 for information.

Click *Cancel* to return to the open document.

Mail

Choose *Mail...* to access your mail system and send each page of recognized text from your currently open document. This command only appears if you have a MAPI-compliant mail system such as Microsoft Mail.

Save

Choose *Save* to write the contents of your current working document to disk. This command is also available as a button in the toolbar.

The Save As dialog box appears when you save a file for the first time. After saving, you can continue working on your document.

Save As

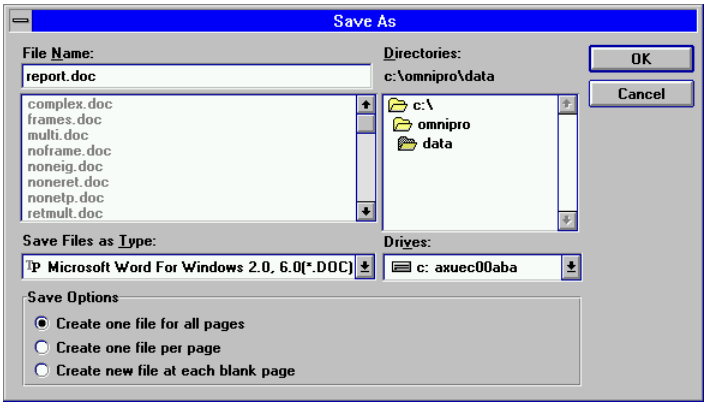
Choose *Save As...* to choose a file format and save a document to disk. This command is also available as a button in the toolbar.

Use this command to save Caere Documents and recognized documents to other file formats.

Saving a File

1 Choose *Save As...* in the File menu.

The Save As dialog box appears.



2 Select a file type in the *Save Files as Type* drop-down list.

See “Supported Output File Formats” on page 238 for a list of supported file formats.

Remember, if you save your image as a Caere Document first, you can reopen and re-edit it, and save it in other file formats as well.

You must perform OCR on any document before you can save it to a text format.

---

3 Type a name for your file in the *File Name* text box.

See the next section for information on how the save option you choose affects the length of the file name.

4 Select a location for your file.

The default location is `omnipro\data`.

5 Select the appropriate option under *Save Options* as described in “Save Options” on page 92.

6 Click *OK*.

OmniPage automatically adds the appropriate file extension to the file name and the current working file returns to the screen.

Click *Cancel* at any time to exit without saving.

Caere Document (*.met)

OmniPage creates a Caere Document the first time you scan a document or open an image. A Caere Document can have up to 256 pages. Each page includes the original image and can vary to include zones and recognized text. When you close the scanned or loaded image, OmniPage prompts you to save the Caere Document. There are advantages to doing this. You can:

- Continue to reopen a Caere Document in OmniPage, make edits, and save it in any other supported file format you wish.
- Use the Verification window to compare recognized text with the original page image.
- Defer recognition.
- Rezone and re-recognize pages at any time.
- Save the time needed to rescan or reload the same page.

You must rescan or reload a document to use it again in OmniPage if you do not save it as a Caere Document.

Saving one or more images as a Caere Document, however, requires more room on the hard drive. The amount depends on the size of the image(s).

Save Options

When you save your document to a file format other than a Caere Document you can select one of three *Save Options.*

Create one file for all pages

Select this to save all the pages in your document as one file. (Blank pages are not saved.) Save the file with a standard file name of eight characters or less.

Create one file per page

Select this to create a separate file for each page in your document and automatically increment file names. (Blank pages are not saved.) Save the file with a file name of five characters or less. OmniPage appends numbers starting with `001`.

For example, if you use `form` as a file name, the first file is named `form001`, the second file `form002`, and so on. The file extension added depends on your choice of file formats: a Word for Windows file would be named `form001.doc`.

Create new file at each blank page

Select this to create a new file after each blank page in your document. (Blank pages are not saved.)

For example, if you want to scan several stacks of pages at once, insert blank pages to separate each batch. OmniPage saves the first stack as one file, detects a blank page, saves the next stack as one file, detects a blank page, and so on.

Save the file with a file name of five characters or less. OmniPage appends numbers starting with `001`.

For example, if you use `form` as a file name, the first file is named `form001`, the second file `form002`, and so on. The file extension added depends on your choice of file formats: a Word for Windows file would be named `form001.doc`.

How Saved Text Appears

The way text appears when you open your recognized document in your target application depends on that application.

For example, if you save a page with text and graphics in ASCII format, only the text will be displayed because ASCII format does not retain graphics. Graphics are only displayed in applications that support graphics.

Normal differences in typeface sizes between applications can result in differences in the page formatting and display of the text. The settings within the application, such as margins, also affect the page layout.

If you use the True Page option (chosen in the OCR settings panel), OmniPage exports text in frames. If your application doesn't accept frames, the text frames are not maintained in their original positions and the text within the frames is displayed in one vertical column.

Applications that support frame-based output have the letters *TP* in front of their names in the *List Files of Type* drop-down list in the Save As dialog box. See Chapter 6, Using True Page, for more information.

Export Image

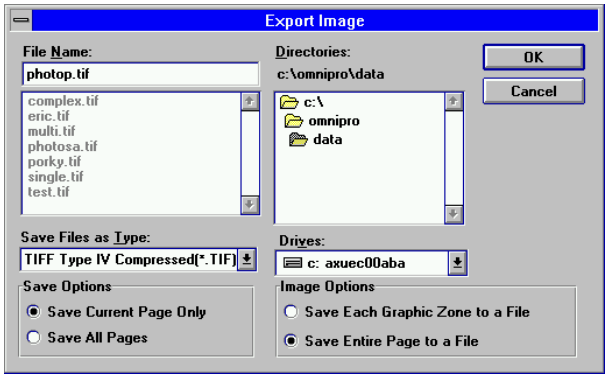
Choose *Export Image...* to save an image to disk in an image file format such as TIFF or PCX. This exports just the original scanned image of a document, not zone or OCR information.

An image file is a "picture" of text and/or graphics. When you open an image file in OmniPage, it appears in the zone window.

Exporting an Image File

You can export an image file after a document has been scanned or loaded.

1 Choose *Export Image...* in the File menu.
The Export Image dialog box appears.



2 Select a file type in the *Save Files as Type* drop-down list.
See “Supported Output File Formats” on page 238 for a list of supported file formats.

3 Type a name for your file in the *File Name* text box.
See “Graphic File Name” on page 96 for information on how the options you choose affect the length of the file name.

4 Select a location for your file.

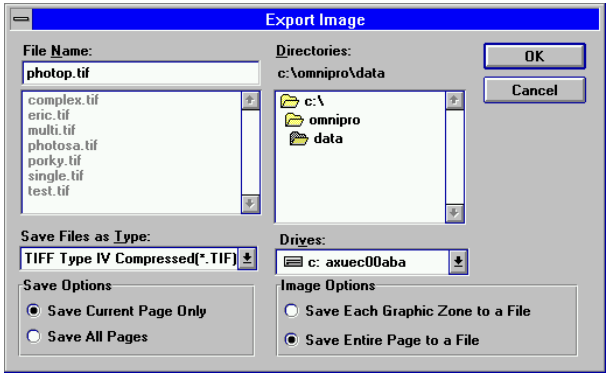
5 The default location is `omnipro\data`.

6 Select *Save* and *Image* options as described in the following sections.

7 Click *OK*.
OmniPage automatically adds the appropriate file extension to the file name. The Export Image dialog box closes and the current working file returns to the screen.
Click *Cancel* at any time to exit without saving.

Save Options

You can select one of two *Save Options.*



- Select *Save Current Page Only* if you want OmniPage to save only the current page image as a file.
- Select *Save All Pages* if you want OmniPage to create a separate file for each page in your document and automatically increment file names starting with `001`.

You must have *Save Page Images in Caere Document* selected in the Preferences settings panel to save an a page to an image file.

Image Options

You can select one of two *Image Options.*

- Select *Save Each Graphic Zone to a File* if you want OmniPage to save only the graphics within your page image. You must create graphic zones on the page image and perform OCR before you can choose this option.
- Select *Save Entire Page to a File* if you want OmniPage to save the entire page image. You do not need to create zones or perform OCR unless you have graphic zones.

You must either choose the *Multiple Columns* zoning option in the Zones settings panel or draw manual zones and identify the graphics as graphic zones to separate graphics from text and export them.

Graphic File Name

The way you match the *Save Options* and *Image Options* affects the length of the file name. The file name `form` is used as an example of how a file would be named in the following combinations of save and image options.

- *Save Current Page Only* and *Save Entire Page to a File:* the file name can have up to eight characters. This creates a one-page image file.

  A PCX file named `form` would be saved as `form.pcx`.

- *Save All Pages* and *Save Entire Page to a File:* the file name can have up to five characters. `00n` is appended, where *n* represents the page number (`001`, `002`, etc.). This creates multiple one-page image files.

  A multiple-page PCX file named `form` would be saved as `form001.pcx, form002.pcx,` and so forth.

- *Save Current Page Only* and *Save Each Graphic Zone to a File:* the file name can have up to seven characters. OmniPage appends a letter to indicate the order of the graphic on the page.

  A PCX file with multiple graphic zones named `form` would be saved as `forma.pcx, formb.pcx,` and so forth.

  This creates one file for each graphic on the current page. Up to 26 files can be created in one directory with this method.

- *Save All Pages* and *Save Each Graphic Zone to a File:* the file name can have up to four characters. OmniPage appends both a number and a letter as an extension.

  A multiple-page PCX file with multiple graphics named `form` would be saved as `form001a.pcx, form001b.pcx,` and so forth. This creates one file for each graphic on every page.

  The number (`00n`) indicates the page number and the letter indicates the order of the graphic on the page. Thus the second graphic on the second page would be named `form002B.pcx`.

Revert to Saved

Choose *Revert to Saved* to undo edits made to a file and return to the last-saved version of the file.

If you accidentally deleted important information in the text window, for example, choose *Revert to Saved* and the file will reappear as it was when you last saved it.

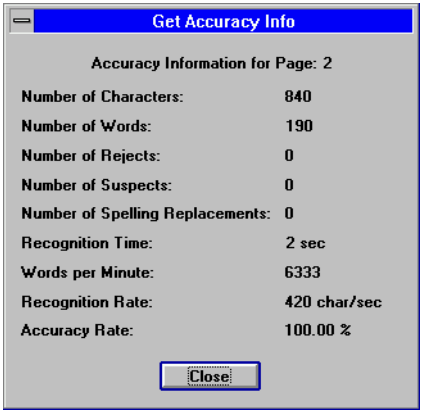
Get Accuracy Info

Choose *Get Accuracy Info...* for a statistical report on recognition accuracy.

Accuracy information is valuable for comparing the effect of different settings on recognition accuracy. For example, if you are not sure about which Scanner settings panel options to choose, you can compare the recognition accuracy percentages of different options.

You can also quickly tell if a poor-quality document is worth editing. If the recognition accuracy rate is less than 97%, it might be quicker to rescan a better copy of the page or to enter the text manually.

The Get Accuracy Info dialog box provides a statistical report for the most recently recognized page.



Number of Characters

This is the number of characters and spaces recognized on the page.

Number of Words

This is the number of words recognized on the page.

Number of Rejects

This is the number of unrecognizable characters. This does not count improper substitutions or incorrectly recognized formatting commands.

Reject characters appear in red in the recognized document. By default, rejects are represented by the tilde (~) character.

Number of Suspects

This is the number of questionable characters that OmniPage made an attempt to recognize. These words are green in the recognized document.

Number of Spelling Replacements

This is the number of words that were corrected automatically by the Language Analyst. These words are blue in the recognized document.

Recognition Time

This is the time it took to break the page down into text and graphics and perform recognition. This does not count scanning time, the time it takes to create zones, or the time spent writing data to disk.

Words per Minute

This is the number of words per minute (wpm) that OmniPage recognized. Assuming that the average word is five characters long, the formula is

`[characters per second ÷ 5] x 60 = wpm`

Recognition Rate

This rate is expressed in characters per second (cps). The formula is

`total number of characters ÷ recognition time = cps`

Accuracy Rate

This is the character recognition accuracy given as a percentage. The formula for Accuracy Rate is

`[number of characters - number of rejects] ÷ number of characters = recognition accuracy`

If the accuracy rate is less than 97%, it might be quicker to rescan a better copy of the page or to enter the text manually.

Save Settings

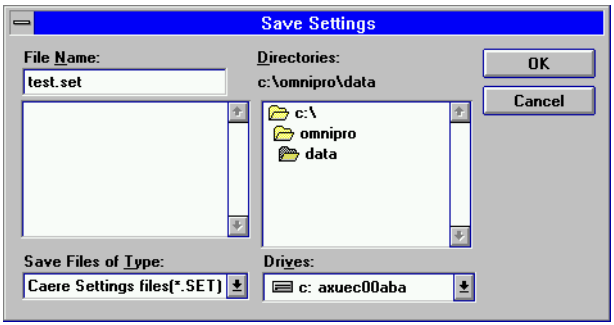
Choose *Save Settings...* to save the currently selected Settings Panel options and language selection(s) (from the Select Languages dialog box) to a settings file (`*.set`) for later use.

Saving settings files is especially useful if you use the same settings often.

Saving Settings

1. Select the Settings Panel options you want to save if they are not set already.
2. Choose *Select Languages...* in the Settings menu.
   Select the language(s) appropriate to your document and click *OK*.

3 Choose *Save Settings...* in the File menu.
The Save Settings dialog box appears.



Caere Settings (.set)* is the only selection in the *Save Files of Type* drop-down list.

4 Type a name for your file in the *File Name* edit box.

5 Select a location for your file.

6 The default location is `omnipro\data`.

7 Click *OK*.

Choose *Load Settings...* in the File menu to load the file. See the next section.

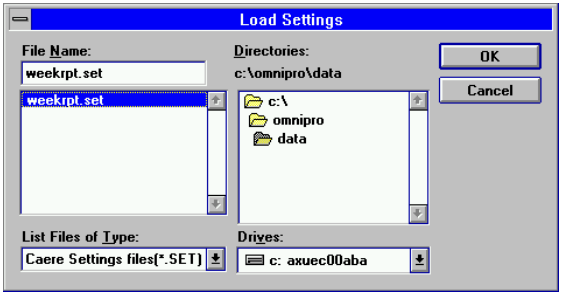
Load Settings

Choose *Load Settings...* to load a previously saved settings file (`*.set`).

A loaded settings file automatically configures the Settings Panel and language selection(s) to preselected values. This is useful for quickly restoring OmniPage to settings required for particular documents.

Loading a Settings File

1 Choose *Load Settings...* in the File menu.
The Load Settings dialog box appears.



Caere Settings (.set)* is the only selection in the *Save Files of Type* drop-down list.

2 Locate and select the settings file to open.

3 Click *OK*.

The settings are loaded immediately into the Settings Panel.

To save a settings file, choose *Save Settings...* in the File menu. See "Deleting *.set, *.trn, *.ud, *.zcn, and *.zon Files" on page 236 for information on how to delete a settings file.

Save Zone Template

Choose *Save Zone Template...* to save manually created zones on a page image as a template.

A zone template file (`*.zon`) is comprised of various zone attributes such as position, order, and zone contents. If you frequently process documents with layouts and content that require the same type of zoning, you can create and save a zone template. Save time by applying it to all documents of the same layout, especially when processing multiple documents.

Automatically drawn zones cannot be saved as a zone template.

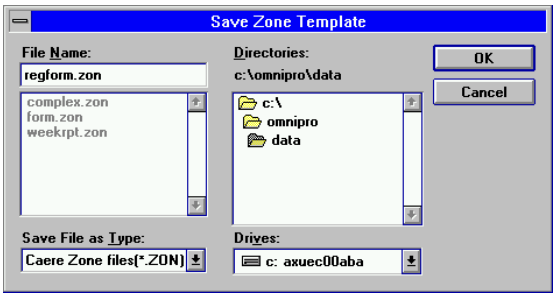
Saving a Zone Template

1 Create manual zones on a page image.

2 See "Manual Zones — Recognize Portions of a Page" on page 48 for an overview of manual zoning.

3 Choose *Save Zone Template...* in the File menu.

The Save Zone Template File dialog box appears.



4 *Caere Zone (`*.zon`)* is the only selection in the *Save Files as Type* drop-down list.

5 Type a name for your file in the *File Name* text box.

6 Select a location for your file.

`Omnipro\data` is the default directory. OmniPage only looks here for zone template files so you must save the file here.

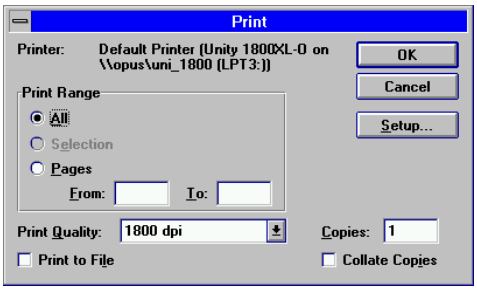
7 Click *OK*.

Saved templates appear in the drop-down list under the Zone button. You can also choose *Use Template...* in the Process menu. This command is described later in this chapter.

Print

Choose *Print...* to print a recognized document. This prints information in the text window but not the zone window. This command is also available as a button in the toolbar.

The dialog box that appears depends on your printer.



Select the desired print options such as print range, print quality, and number of copies. Click *OK* to start the print job. Click *Cancel* to exit without printing or saving the selected print options.

Publish to Envoy

Choose *Publish to Envoy...* to save recognized text and any retained graphics as a Novell Envoy runtime file.

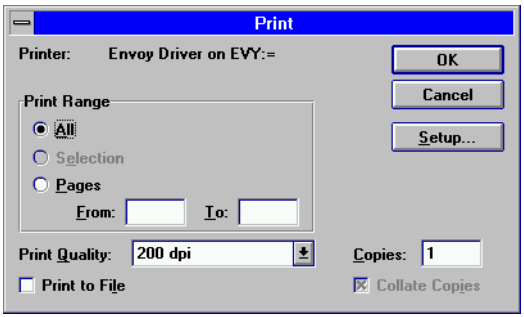
An Envoy file displays information on your screen as if it were printed. Envoy preserves your document's fonts and page layout. Text and graphics appear as they did in the OmniPage text window.

You cannot edit an Envoy file as you would edit a file in a word processor, but you can rearrange, combine, and delete whole pages of the file. You can annotate an Envoy file, print it, and copy and paste from an Envoy file into other applications. You can send the file to other users and they will be able to open it on their system. Envoy files are a self-opening; you do not need to install either OmniPage or Envoy to view an Envoy file. See "Opening a Novell Envoy Runtime File" on page 103.

Saving Recognized Text as a Novell Envoy Runtime File

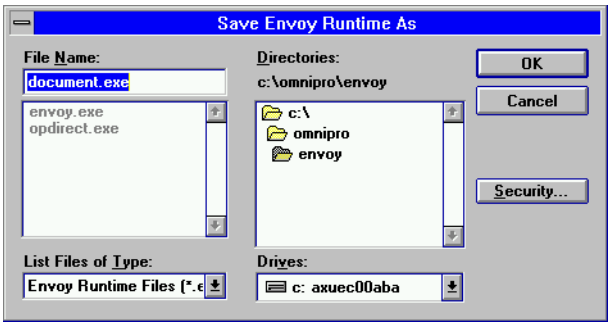
1 Choose *Publish to Envoy...* in the File menu.

The Print dialog box appears. The Envoy driver is selected automatically.



2 Click *OK*.

The Save Envoy Runtime As dialog box appears.



Envoy Runtime Files (.exe)* is the selection in the *Save Files as Type* drop-down list.

3 Type a name for your file in the *File Name* text box.

4 Select a location for your file.

The default directory is `omnipro\envoy`.

5 Click *OK*.

Your file will automatically be saved as an Envoy Runtime file with an *.`exe` extension.

Opening a Novell Envoy Runtime File

An Envoy file (`*.exe`) is self-opening: it includes a scaled-down version of the Novell Envoy application. You can open the file on the same kind of computer it was created on (Macintosh or PC) even without Envoy or OmniPage installed.

To open your Envoy file, double-click its file name (`*.exe`) in the Windows File Manager. Your file will open in the Novell Envoy Runtime Viewer with the title "Embedded Document."

The file will open in the regular Envoy viewer if you have the complete version of Envoy installed. You can print, annotate, and share this file on any the system.

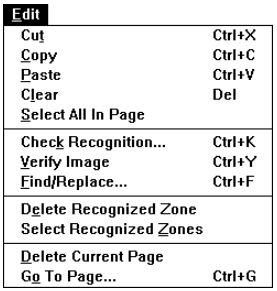
Descriptions appear on the Envoy title bar at the top of the screen and on the status bar at the bottom of the screen to define what a selected button or command does and actions you can do next as you perform a task.

You cannot import or open any file other than the file that is attached to the Envoy runtime viewer.

Exit

Choose *Exit* to quit the OmniPage program. If the current working document has changed since the last save, a prompt appears asking if you want to save the document. Click *Yes* to save changes. Click *Cancel* to return to the open document.

The Edit Menu



The Edit menu lets you revise text in the text window and work with images in the zone window. Edit menu commands include:

- Cut
- Copy
- Paste
- Clear / Clear All Zones
- Select All in Page
- Check Recognition
- Verify Image
- Find / Replace
- Delete Recognized Zone
- Select Recognized Zones
- Delete Current Page
- Go to Page

Cut

Choose *Cut* to delete selected text from the recognized document and copy it to the Clipboard. This command is also available as a button in the toolbar.

1 Select the text you want to cut.

2 Choose *Cut* or click the Cut button.
The selected text disappears and is copied to the Clipboard.

Text stored on the Clipboard may be pasted anywhere (except into a graphic) in the document. The text remains on the Clipboard until new text is cut or copied.

Copy

Choose *Copy* to copy selected text in a recognized document to the Clipboard. This command is also available as a button in the toolbar.

1 Select the text you want to copy.

2 Choose *Copy* or click the Copy button.
The selected text remains on the page as it is but is also copied to the Clipboard.

Text copied to the Clipboard may be pasted anywhere (except into a graphic) in a document. The text remains on the Clipboard until new text is cut or copied.

Paste

Choose *Paste* to place cut or copied text from the Clipboard into the recognized document. This command is also available as a button in the toolbar. See the preceding two sections if you do not know how to cut or copy text.

Pasted text appears at the cursor location. A copy of the pasted text remains on the Clipboard until new text is cut or copied.

Clear

Choose *Clear* to delete selected recognized text permanently.

1 Select the text you want to clear.

2 Choose *Clear*.
Cleared text is not copied to the Clipboard.

Clear All Zones

Choose *Clear All Zones* to delete all zones on a page image in the zone window. This command is also available as a button in the toolbar.

Clear All Zones appears in the menu only when the zone window is active. When you clear zones, only the zone borders are deleted; the image itself remains the same. After the zones are cleared, you can create new zones manually, automatically, or by using a zone template.

Select All in Page

Choose *Select All in Page* to automatically select the entire contents of a recognized page in the text window. This command is available only when the text window is active.

To deselect selected text, click anywhere within it or choose *Select All in Page* again.

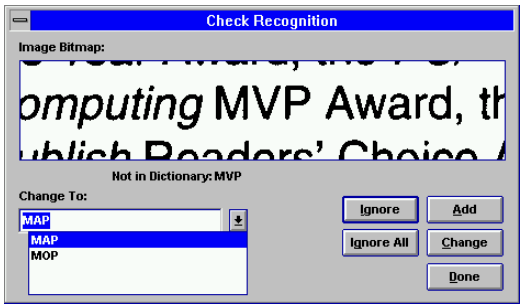
Check Recognition

Choose *Check Recognition...* to check for errors in a recognized document. This command is also available as a button in the toolbar.

OmniPage uses the currently selected main and user dictionaries and language(s) to check recognition. The Check Recognition operation stops at:

- Blue words: these were replaced by the Language Analyst.
- Green words: these have questionable characters that OmniPage made an attempt to recognize.
- Red text: these are unrecognizable characters in a word that are replaced with a red reject character (~ is the default).
- Words not found in the main or user dictionary.

When OmniPage finds a suspect word, the Check Recognition dialog box shows the original image of the word as it appeared in the original page.



You can only see the original image if the original page images are saved in the Caere Document. Select *Save Page Images in Caere Document* in the Preferences setting panel before loading an image or scanning a document.

Choose one of the following options for a word flagged as a suspect. After you choose an option for the word, OmniPage automatically finds the next suspect.

Ignore

Click this to allow a word to remain as is and go on to find the next suspect.

Ignore All

Click this to ignore all instances of a word in the document and go on to find the next suspect.

Add

Click this to add a word to the current User Dictionary and go on to find the next error. Other occurrences of the word in the current document will be checked if the word was a "suspect" (green). OmniPage accepts future occurrences of the word when you use the same user dictionary for future documents.

Change

Click this to replace a word with the word in the *Change To* edit box. To place a word in the *Change To* edit box, you can either type in a word or select a word from the drop-down list.

The original text of a word corrected by the Language Analyst appears as the first word in the list box in case you want to change it back.

Done

Click this to exit the Check Recognition operation. Any changes made up to this point will be retained. Choose *Save* in the File menu to save the changes to the document.

Verify Image

Choose *Verify Image* to view the original image of recognized text in the Verification window. The Verification window shows a clear close-up of the original scanned word image and surrounding area of selected text.

This lets you double-check any questionable words in the text window against its original scanned image.

In order to verify images, the original page images must be saved along with the Caere Document. Select *Save Page Images in Caere Document* in the Preferences settings panel before loading an image or scanning a document.

Saving the original page images slows down processing slightly and uses more disk space.

Verifying an Image

1. Double-click any word in the text window or select a word and choose *Verify Image* in the Edit menu.

   The Verification window appears showing the original image of the selected area of text.

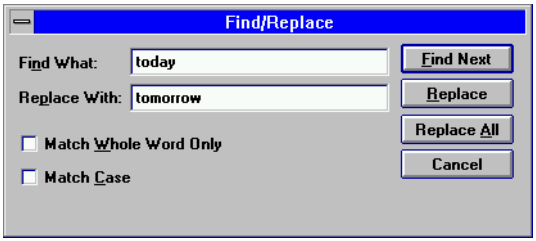
   You can retype the selected word while the Verification window is open if you need to.

2. Click anywhere outside the Verification window to close it.

The Verification window may not show an accurate image for text that is cut and pasted from one page to another.

Find/Replace

Choose *Find/Replace* to find a word or set of characters in the recognized document and replace it with another word if desired. This command is also available as a button in the toolbar.



By default, when you search for a word, all occurrences of letter combinations that match the word are found. For example, if *jelly* is the search word, OmniPage would find the *jelly* in *jellyfish.*

Finding a Word

1. Type the search word or character(s) in the *Find What* edit box.
2. Select specific search options, if desired.

- Select *Match Whole Word Only* to find only the words that exactly match the length of the search word. Compound words that contain the search word within them will not be found. For example, if you select this option and search for *light*, OmniPage will not flag *lightbulb*.
- Select *Match Case* to find only the words that exactly match the upper- and lower-case attributes of the search word. For example, if you select this option and search for *light*, OmniPage will not flag *Light*.

3 Click *Find Next*.

The first occurrence of the word is highlighted.

4 To continue searching, click *Find Next* again.

5 Click *Cancel* to close the dialog box when you are done.

Replacing a Word

1 Type the word or set of characters that you want to replace in the *Find What* edit box.

2 Type the desired replacement word in the *Replace With* edit box.

3 Select specific search options if desired.

- Select *Match Whole Word Only* to find only the words that exactly match the length of the search word. Compound words that contain the search word within them will not be found. For example, if you select this option and search for *light*, OmniPage will not replace *lightbulb*.
- Select *Match Case* to find only the words that exactly match the upper- and lower-case attributes of the search word. For example, if you select this option and search for *light*, OmniPage will not replace *Light*.

4 Click *Find Next*.

The first occurrence of the word is highlighted.

5 Click *Replace* to insert the replacement word.

OmniPage automatically looks for the next occurrence of the search word.

You can click *Replace All* to replace all instances of a word.

6 Click *Cancel* to close the dialog box when you are done.

Delete Recognized Zone

Choose *Delete Recognized Zone* to delete a selected text or graphic zone in a recognized page. Each zone is surrounded by a frame. This command is available in the Edit menu when the text window is active.

Delete a Text Zone

1 Place your cursor in the zone to delete.

2 Choose *Delete Recognized Zone* in the Edit menu.
The zone and all text in it disappears.

Delete a Graphic Zone

1 Choose *Select Recognized Zones* in the Edit menu.
All zones on in the text window are selected. The text zones have handles on the frames.

2 Click in the graphic zone to delete.
The zone you clicked in is deleted but other zones are not.
You can also delete a text zone this way.

See the next section for more information on selecting zones.

Select Recognized Zones

Choose *Select Recognized Zones* to select all text and graphic zones on a recognized page in the text window. This command only appears when the text window is active. A check mark appears next to this command when it is selected.

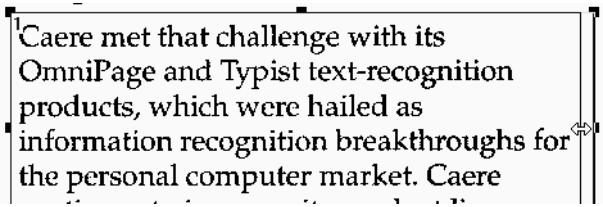
You can resize or reposition selected zones to change the page layout. Each zone is surrounded by a frame. When you select zones, handles appear on the frames.

Place your cursor inside a zone, hold down the Alt key, and click the right mouse button to select that zone individually.

Resize a Zone

1 Hold your cursor over a zone handle until it turns into a two-way arrow.

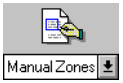
2 Hold down the mouse button and drag to resize the zone.



Move a Zone

1 Place your cursor inside the zone until it turns into a four-way arrow.

2 Hold down the mouse button and drag the zone to the desired location.



To deselect zones, choose *Select Recognized Zone*s in the Edit menu again. The check mark disappears and the zones are deselected.

Delete Current Page

Choose *Delete Current Page* to delete the page being viewed. You may want to delete a page in your document that was poorly scanned or recognized. A dialog box prompts you to confirm the selection if *Prompt Before Deleting Pages* is selected in the Preferences settings panel.

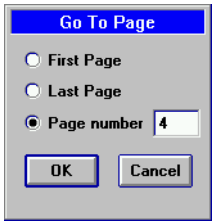
When you delete a page, everything is discarded including the page image and recognized text.

A multiple-page document opens to the next page after you delete a page. A single-page document closes after its only page is deleted. A dialog box warns you first that you are deleting the only page in your document and that it will close if you continue. The page and therefore your one-page document is deleted if you continue.

Go to Page

Choose *Go To Page...* to open the Go To Page dialog box and switch to another page in the current document. Both the text window and zone window display the selected page.

In the Go To Page dialog box, you can select either *First Page* or *Last Page*, or type a page number in the *Page Number* edit box.



Click *OK* to go to the selected page. Click *Cancel* to exit and return to the current page.

OmniPage opens the last page if you type a number greater than the actual number of pages. If you accidentally type a non-number, such as a letter, OmniPage does nothing.

There is also a page number button with arrows in the status bar at the bottom of the page. Click the left arrow to move one page backward and the right arrow to move one page forward in the document. Click the current page number button to open the Go To Page dialog box.

The Format Menu

The Format menu lets you format character and paragraph attributes while you edit a recognized document in the text window. Format menu commands include:

- Font
- Paragraph

Files saved in ASCII or ANSI format do not retain any formatting other than spaces and carriage returns.

Font

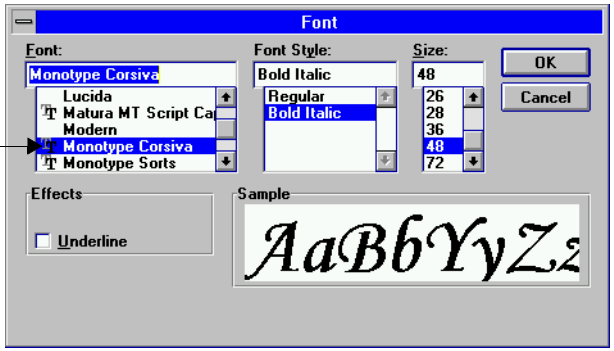
Choose *Font...* to change the attributes of a selected character or section of text in a recognized document. Fonts options are also available as buttons in the toolbar as described later in this section.

Applying Character Formatting

1 Select the text you want to format.

2 Choose *Font...* in the Format menu.
The Font dialog box appears.

TrueType fonts have a *TT* in front of their names.



3 Make the desired formatting selections in the Font dialog box.
TrueType fonts have the letters *TT* in front of their names. The *Sample* box illustrates the selected attributes.

4 Click *OK*.
The selected text changes accordingly.
Click *Cancel* to exit without applying formatting selections.

Font

Select a font in the *Font* list box. You can type a letter in the *Font* edit box to skip to the fonts beginning with that letter.

Font Style

Select *Regular* to return selected characters to an unformatted state. Bold, italic, and underlined characteristics disappear.

Select *Italic* to change selected characters to an italicized format.

Select *Bold* to change selected characters to a boldfaced format.

Select *Bold Italic* to change selected characters to a boldfaced, italicized format.

Size

Select from a range of font sizes in the *Size* list box.

Effects

Select *Underline* to change selected characters to an underlined format.

Font Style Buttons

You can also use the Bold, Italics, and Underline buttons in the text window to format selected text.

Click the Bold button to change selected characters to a boldfaced format.

Click the Italic button to change selected characters to an italicized format.

Click the Underline button to change selected text to an underlined format.

Paragraph

Choose *Paragraph...* to change the attributes of a selected paragraph in a recognized document. You can also use the formatting buttons in the text window to format selected paragraphs.

Applying Paragraph Formatting

1 Select one or more paragraphs.

You can place your cursor in a paragraph if you want to select just one.

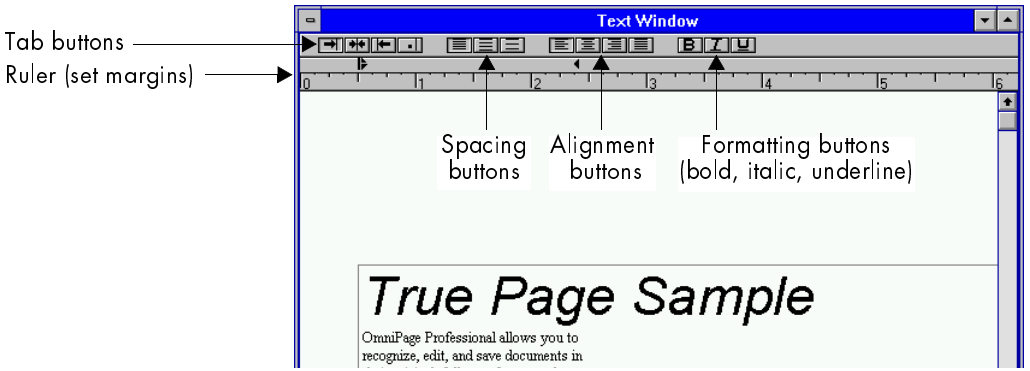
2 Choose *Paragraph...* in the Format menu.
The Paragraph dialog box appears.

3 Select an option in the *Spacing* drop-down list.
- Select *Single* for single-spaced lines.
- Select *Double* for double-spaced lines.
- Select *Triple* for triple-spaced lines.

4 Select an option in the *Alignment* drop-down list.
- Select *Left* for left-aligned text.
- Select *Center* for center-aligned text.
- Select *Right* for right-aligned text.
- Select *Justify* for justified text.

5 Click *OK* to format the selected paragraph(s).

You can also use the buttons in the text window to format paragraphs as described in the following sections.

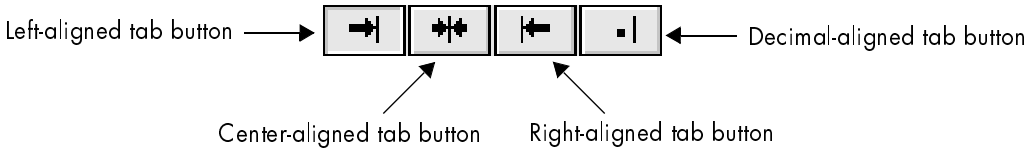


Tab Buttons

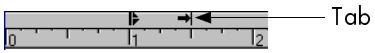
You can use the Tab-setting buttons in the text window to insert tabs in your recognized document.

1 Select one or more paragraphs.
You can place your cursor in a paragraph if you want to format just one.

2 Click the appropriate Tab-setting button (left, center, right, or decimal).



3 Click the area in the upper half of the ruler where you want to place the tab stop.



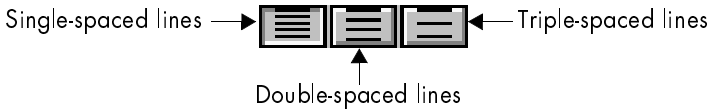
4 Repeat steps 1 through 3 to continue setting tabs.

Line-spacing Buttons

You can also use the Line-spacing buttons in the text window to set line spacing for selected paragraphs.

1 Select one or more paragraphs.
You can place your cursor in a paragraph if you want to format just one.

2 Click the appropriate Line-spacing button.

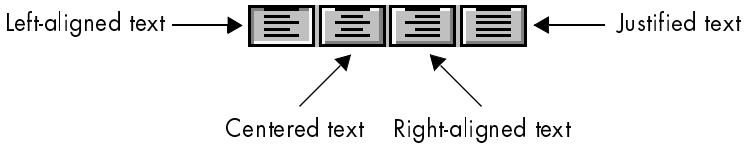


Alignment Buttons

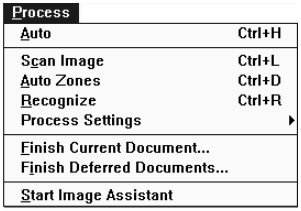
You can use the Alignment buttons in the text window to align selected paragraphs.

1 Select one or more paragraphs.
You can place your cursor in a paragraph if you want to format just one.

2 Click the appropriate Alignment button.



The Process Menu



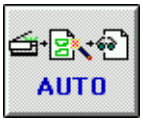
The Process menu lets you perform each step of the OCR process as well as other OCR-related functions. Process menu commands include:

- Auto / Stop
- Scan Image / Load Image
- Auto Zones / Manual Zones / Use Template
- Perform OCR / Defer OCR / Train OCR
- Process Settings
- Finish Current Document
- Finish Deferred Documents

Some of the Process menu commands are available as buttons in the toolbar. The Process Settings commands are available in the Image, Zone, and OCR button drop-down lists. The Process Settings commands change according to the currently selected button commands and vice versa.

See Chapter 5, Direct Input, for information on Process menu commands that appear only in Direct Input mode.

Auto



Choose *Auto* to automatically start and finish processing each page of a new document or finish processing the current page of an open document. This command performs the same function as the *AUTO* button in the toolbar.

Automatic processing of a document is determined by the currently selected *Process Setting* commands. For example, if *Scan Image*, *Auto Zones*, and *Perform OCR* are selected as the processing commands, the following process occurs when you choose *Auto*:

1. OmniPage scans the page in your scanner and displays the image in the zone window.
2. OmniPage creates zones on the page image according to the zoning method selected in the Settings Panel.
3. OmniPage performs OCR on the page image and displays the recognized text in the text window.

   This same process repeats for every page in a multiple-page document.

How OmniPage scans, zones, and recognizes a page depends on the options chosen in the Settings Panel.

When a document is already open to an unfinished page image, you can choose *Auto* to finish processing that page according to the selected processing commands. For example, if your document has been scanned but not zoned, choose *Auto* to create zones and recognize the page according to the selected zone and OCR commands.

As automatic processing begins, the *Auto* command changes to *Stop*. Choose *Stop* to halt processing at any time.

Stop

Choose *STOP* if you want to discontinue processing at any time. This command performs the same function as the *STOP* button in the toolbar.

For example, you may want to stop scanning your page if you realize that the page is misaligned or that the wrong zoning method was set.

Scan Image

Choose *Scan Image* to scan a page in your scanner. This command performs the same function as the Image button when *Scan Image* is selected in the drop-down list.

This command only appears in the menu if you have installed the Scan Manager. Select your default scanner in the Scan Manager before scanning (see “Scan Manager Installation” on page 8). Select the appropriate Scanner options in the Settings Panel as well.

Use your right mouse button to click the Image button and automatically open the Settings Panel to Scanner options.

Scanning a Page

Select Scanner options in the Settings Panel before scanning. A progress meter appears during a scan and the status bar displays progress. To cancel scanning, click the *STOP* button in the toolbar. When scanning is done, the image appears in the zone window.

Scanning Multiple Pages

You can scan pages one at a time to create a multiple-page Caere Document in OmniPage or scan a stack of pages if you have an automatic document feeder (ADF). A Caere Document can have 256 pages.

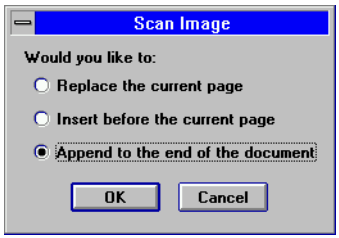
Click the Image button or *Auto* after a document is scanned to scan or load another page and add it to the current open document. See the next section, "Adding a Page to a Scanned Image," for information on adding pages.

Select *Scan Until Empty* in the Scanner settings panel if you have an ADF and want to scan a stack of pages. Select the desired processing commands and choose *Auto* to begin automatic processing. The pages are scanned and processed in the order that they are placed in the scanner and combined into one working document.

Adding a Page to a Scanned Image

A scanned image becomes your working document. If a document is already open when you begin to scan, the newly scanned image can be added to it.

- The newly scanned page automatically becomes page two of a one-page document.
- The newly scanned page automatically is added to the last page of a document if it is open to the last page.
- The Scan Image dialog box appears if the open document has multiple pages and is not open to the last page. Choose an option and click *OK*.



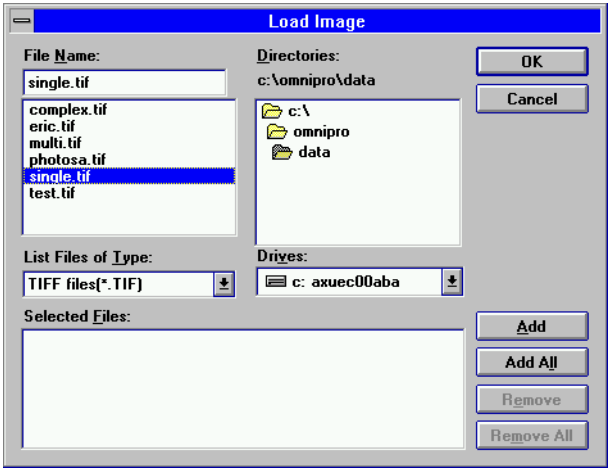
Load Image

Choose *Load Image* to open a previously saved image file. This command performs the same function as the Image button when *Load Image* is selected in the drop-down list.

An image file is an electronic "picture" of text and / or graphics that has been saved in an image file format such as TIFF or PCX. When you load an image file in OmniPage, it appears in the zone window. See "Supported Input File Formats" on page 239 for a list of supported image file formats.

Loading an Image File

1 Choose *Load Image* in the *Process Settings* cascading menu.
The Load Image dialog box appears.



2 Select a file type in the *List Files of Type* drop-down list.
Files of that type appear under the *File Name* list box.

3 Locate and select the file to load.

4 Click *Add*.
The file appears in the *Selected Files* list box.

- Click *Add All* to add all files in the selected directory.
- Select a file and click *Remove* to remove it from the *Selected Files* list box.
- Click *Remove All* to remove all files from the *Selected Files* list box.
- Click *Cancel* to exit without loading an image file.

5 Click *OK*.
The file opens in the zone window. For a multiple-page image file, you must click the Image button to load each consecutive page in the file.

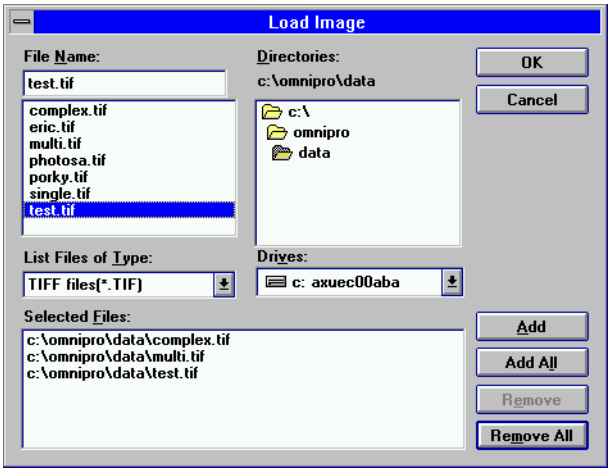
Loading Multiple Image Files

You can load multiple image files all at once if you use automatic processing. Select *Load Image*, *Auto Zones*, and *Perform OCR* in each process button drop-down list.

1 Choose *Auto* in the *Process Settings* cascading menu.
The Load Image dialog box appears.
See the previous section for information on the dialog box buttons.

2 Select a file type in the *List Files of Type* drop-down list.
Files of that type appear under the *File Name* list box.

3 Select a file to load and click *Add*.
The file appears in the *Selected Files* list box. You can select up to 250 files.



4 Repeat for each file you want to load.

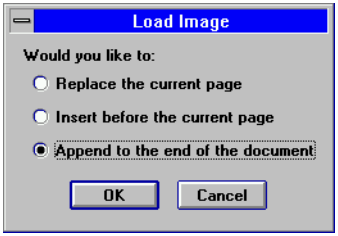
5 Click *OK* when you have selected all the files to load.
OmniPage loads, zones, and recognizes each file in the order it was selected.
If you had selected a non-automatic processing option, such as Manual Zones, you would draw zones and click *AUTO* to proceed with OCR. OmniPage would then automatically open the Load Image dialog box. Click OK to continue processing.

Adding a Page to a Loaded Image

A loaded image becomes your working document. If a document is already open when you choose a file to load, the newly loaded image is added to it.

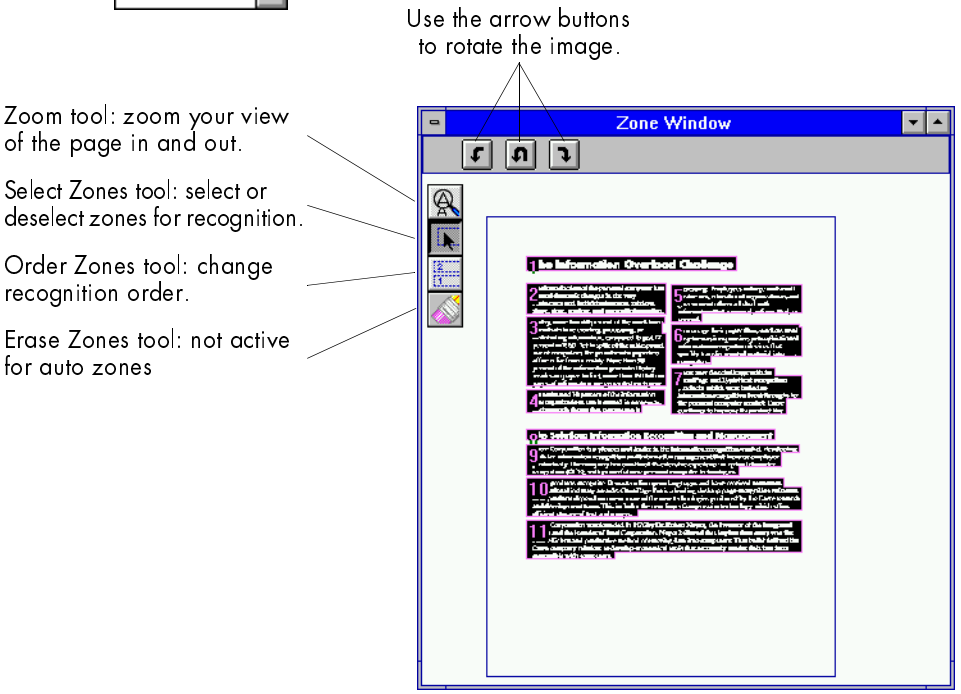
- The new document automatically becomes page two of a one-page document.
- The new document automatically is added to the last page of a document if it is open to the last page.

- The Load Image dialog box appears if the open document has multiple pages and is not open to the last page. Choose an option and click *OK*.



Auto Zones

Choose *Auto Zones* to have OmniPage automatically draw and order zones on a scanned or loaded page image in the zone window. Each zone has a number indicating its recognition order.



This command performs the same function as the Zone button when *Auto Zones* is selected in the drop-down list.

OmniPage uses the selected Zones option in the Settings Panel when zoning: *Multiple Columns, Single Column or Table,* or *One Zone.* For more information about each of these options, see "Zones Options" on page 163.

You are prompted to delete the current zones before auto zoning if a page already has zones. Click *Yes* to proceed.

Each zone has a number indicating its recognition order. Use the automatic zone tools to reorder or deselect zones.

Reordering Zones

Zone order determines the order in which text is recognized and positioned on your recognized document. You can reorder zones on a zoned page.

1. Click the Order Zones tool.
   The zone numbers disappear.
2. Click in a zone.
   The number 1 appears in the zone. It will be recognized first.
3. Click in another zone.
   The number 2 appears in the zone. It will be recognized second.
4. Continue until all the zones are in proper recognition order, and then choose the appropriate OCR command.

OmniPage numbers the zones in successive order if you do not reassign numbers to them.

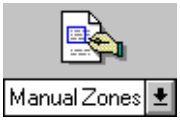
Deselecting Zones

You can deselect zones on an image in the zone window. That portion of the page will not be recognized during OCR.

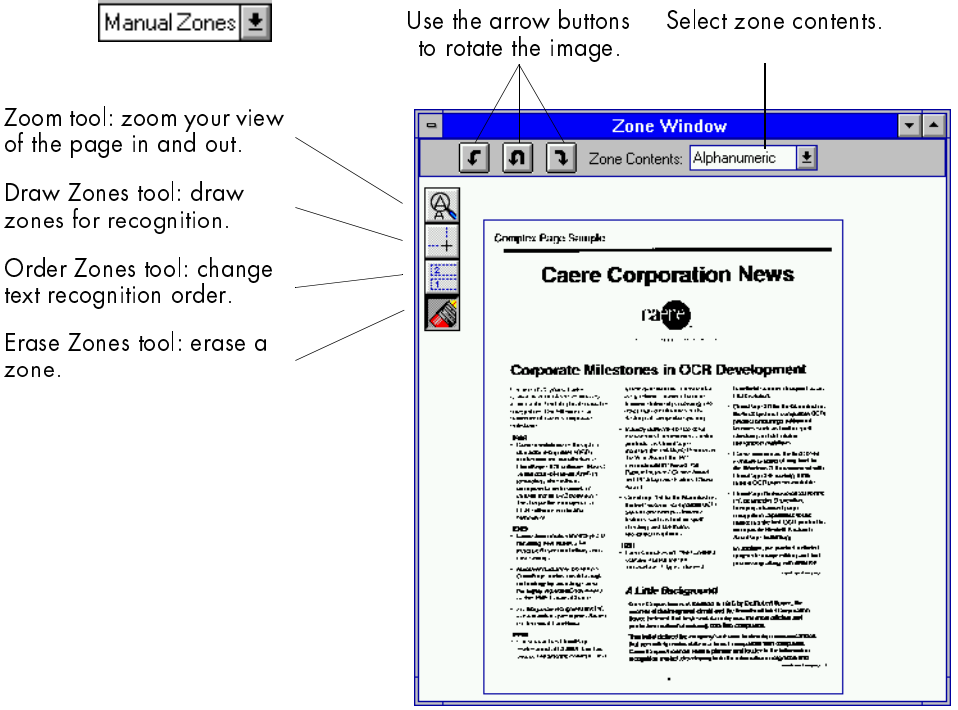
1. Click the Select Zones tool.
2. Click each zone you want to deselect.
   A zone changes from black to white when it is deselected. It will not be recognized. To reselect a zone, click it again with the Select Zones tool.

   You can click the Clear All Zones button in the toolbar to delete all zones on the image.
3. Choose the appropriate OCR command when you are done.

Manual Zones



Choose *Manual Zones* to draw, order, and specify zones manually on a scanned or loaded page image in the zone window.



This command performs the same function as the Zone button when *Manual Zones* is selected in the drop-down list.

OmniPage uses the selected Zones option in the Settings Panel when zoning: *Multiple Columns*, *Single Column or Table*, or *One Zone*. For more information about each of these options, see "Zones Options" on page 163.

You are prompted to delete the current zones before zoning if a page already has zones. Click *Yes* to proceed.

Use the manual zone tools to draw, reorder, resize, move, or erase zones. You can also specify a zone's contents: graphic, numeric, and so forth.

Drawing Zones

1 Click the Draw Zones tool.

2 Enclose an area you want as a zone by holding the mouse button down and dragging the mouse.

When you have enclosed the desired area, release the mouse button.

3 Continue using the mouse to draw zones in the page image until you have finished.

You can draw up to 64 separate zones of which 26 can be graphic zones. Any area of the image that is not part of a zone will not be recognized.

A number appears in each zone indicating the order in which the zone will be recognized. To reorder zones, use the Order Zones tool.

Resizing Zones

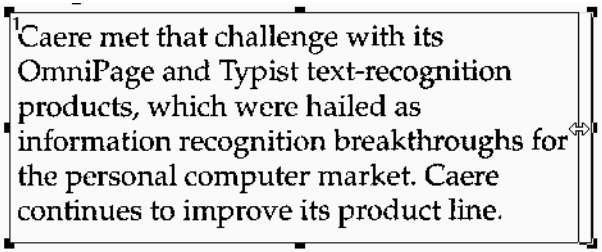
1 Click the Draw Zones tool.

2 Click a zone to select it.

Handles appear on the zone border.

3 Hold your cursor over a handle so that it turns into a two-way arrow.

4 Hold the mouse button down, and drag the mouse to enlarge or reduce the zone as necessary.



Reordering Zones

Your page must already have zones on it to reorder.

1 Click the Order Zones tool.

The zone numbers disappear.

2 Click in the zone you want to recognize first.

The number 1 appears in the zone.

3 Click in the next zone you want recognized second.

The number 2 appears in the zone.

4 Continue until all the zones are in proper recognition order, and then choose the appropriate OCR option.

Moving Zones

1 Click the Draw Zones tool.

2 Click a zone to select it.

3 Handles appear on the zone.

4 Place your cursor inside a zone so that it changes to a four-way arrow.

5 Hold down the mouse button, and drag the zone wherever you want it.

Only the zone borders can be moved; the contents of the page image remain in the same place.

Erasing Zones

Your page must already have zones on it to erase.

1 Click the Erase Zones tool.

2 Click each zone you want to erase.

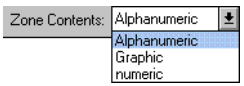
The zone border disappears but the text within in it does not. This portion of text will not be recognized.

Assigning Zone Contents Files

You can assign zone contents files to manually drawn zones for better recognition accuracy. For example, if your image has a text paragraph followed by a table or numbers, you can draw separate zones around each and assign an alphanumeric zone contents file to the paragraph and a numeric zone contents file to the table.

The graphic zone contents file is especially useful for identifying graphics. See for a tutorial exercise that illustrates this.

To assign zone contents files to zones:



1 After drawing zones manually, click within a zone to select it.

2 Select the appropriate zone contents file from the *Zone Contents* drop-down list.

3 Repeat steps one and two for any other zones you wish.

You can change a zone contents assignment at any time before recognition.

You can also create your own zone contents file for specific characters. See “Edit Zone Contents File” on page 146 for information.

Use Template

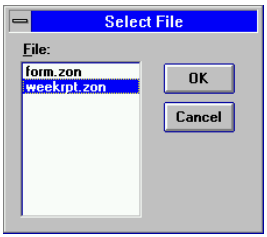
Choose *Use Template...* to create zones by applying a previously created zone template file (`*.zon`). Use the *Save Zone Template...* command in the File menu to create a zone template. See “Save Zone Template” on page 100 for information.

A zone template file is composed of various zone attributes such as position, order, and contents. If you frequently process documents with layouts and content that require the same type of zoning, apply a zone template to save processing time.

Applying a Zone Template

1 Choose *Use Template...* in the Process menu.

The Select File dialog box appears. It lists all zone template files in the `omnipro\data` directory, the default location for these files.



2 Select the zone template to use for the current page image.

3 Click *OK*.

OmniPage draws zones on the page image according to the selected zone template.

Click *Cancel* to exit without applying the zone template.

You can also select a zone template directly from the Zone button drop-down list. The same zone template is applied to each page if you use auto processing. See “Deleting *.set, *.trn, *.ud, *.zcn, and *.zon Files” on page 236 for information on how to delete a zone template.

Perform OCR

Choose *Perform OCR* to recognize text on the current page. This command performs the same function as the OCR button when *Perform OCR* is selected in the drop-down list.

Before performing OCR, make sure the appropriate OCR options are selected in the Settings Panel.

Use your right mouse button to click the OCR button when it is available and automatically open the Settings Panel to OCR options.

If there are no zones on the page when you select *Perform OCR*, OmniPage automatically creates zones according to the selected Zone command. If *Manual Zones* is currently selected, OmniPage ignores this and draws zones automatically.

Defer OCR

Choose *Defer OCR* to delay text recognition of one or more pages of a document during automatic processing. This command performs the same function as the OCR button when *Defer OCR* is selected in the drop-down list. OCR can be a time- and memory-intensive process so you may want it to take place while you are away from your computer.

For example, you can load or scan multiple documents, zone them, and then delay OCR to a more convenient time. At this point, you have two choices:

- Leave the document open and finish it later.

  You would choose *Finish Current Document...* in the Process menu to finish processing an open document. The command is described further in this chapter.

- Save the document as a Caere Document or as an image file, close it, and finish it later.

  You would choose *Finish Deferred Documents...* in the Process menu to open deferred documents, and specify an OCR time and other recognition options. The command is described further in this chapter.

Train OCR

Choose *Train OCR* to create a character training file (`*.trn`) that assists OmniPage during text recognition and allows better recognition of special characters. This command performs the same function as the *Train OCR* command under the OCR button.

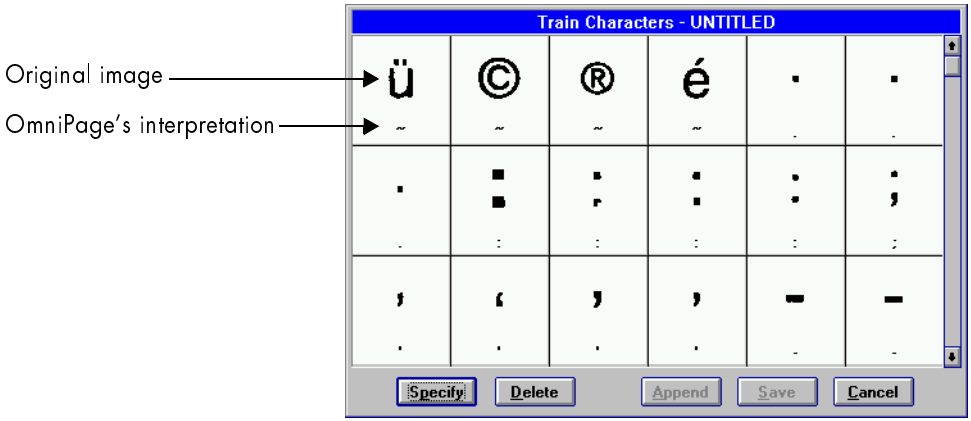
A character training file is a set of pre-recognized text characters that you create. OmniPage compares the file with the characters in the page image during recognition. This helps identify unusual characters such as copyright symbols.

Creating a Character Training File

1. Open an image file or scan an image that includes the characters you want to train.
2. Create manual or automatic zones on the document.
3. Choose *Train OCR* in the Process Settings cascading menu.

   You can also choose the command in the OCR button drop-down list and then click the OCR button.

   OmniPage makes three recognition passes over the document and the Train Characters dialog box appears.



Which Characters to Specify

OmniPage places suspect characters and those it could not identify at the top of the dialog box. Beneath each character is a tilde or OmniPage's attempted identification of that character. A tilde (the default Reject Character) means OmniPage could not identify the character.

Characters OmniPage believes it identified correctly are listed alphabetically below the suspect characters.

You do not have to train OmniPage to recognize every character in the dialog box. OmniPage's word-comparison ability enables it to identify most words correctly, even if a character within the word is unrecognizable. Select *Use Language Analyst* in the OCR settings panel to further improve word recognition.

Look for:

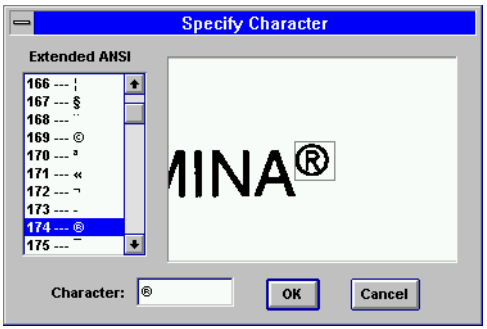
- uncommon characters such as the copyright symbol (©)
- ligatures (fl and ff, for example) because the run-together text could cause identification problems
- common errors, such as a 5 being recognized as an S
- characters in a specialized font, such as *Zapf Chancery*

The next section explains how to identify, or *specify*, characters.

Specify

1 Select a character and click *Specify*, or double-click the character.

The Specify Character dialog box appears. It shows how the selected character appeared in the original page image.

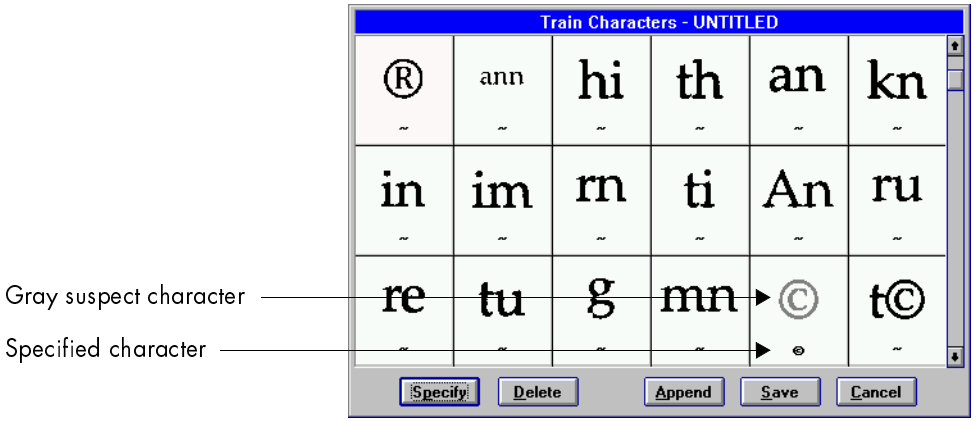


2 Look for the character in the *Extended ANSI* list box at the left.

- Select the character if you find it.
  It appears in the *Character* text box.
- Type the character in the *Character* text box if it is not in the list.
  You will not be able to type all characters in the text box. However, you can cut and paste characters from other sources if you have them, or use Alt-number keyboard codes.

3 Click *OK*.

The specified character now appears below the suspect character.



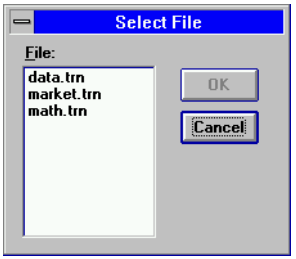
The symbol turns gray to indicate that a character has been specified for it. OmniPage will identify this character correctly when it recognizes your document.

4 Specify as many characters as necessary.
You can select a specified character and click *Delete* to remove it from the Train Characters dialog box.

5 Click either *Append* or *Save* when you are done.
See the following sections for information on these two options.

Append

1 Click *Append* to add the current set of trained characters to another training file.
The Select File dialog box appears.



It displays a list of existing character training files.

2 Select a file.

3 Click *OK*.

The character set you just trained is appended to the selected file.

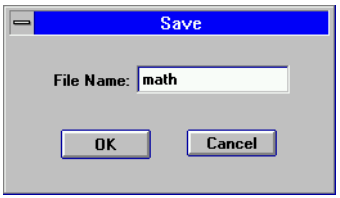
A dialog box asks if you want to make this the current training file and recognize your document.

- Click *Yes* to recognize your page image and apply the training file you just created.
- Click *No* if you want to return to the OmniPage screen without recognizing the image.

Save

1 Click *Save* to save the trained characters to a new file.

The Save dialog box appears.



2 Enter a name in the *File Name* text box.

3 Click *OK*.

Training files are saved by default to the `omnipro\data` directory. This is the only place OmniPage looks for the files.

A dialog box asks if you want to make this the current training file and recognize your document.

- Click *Yes* to recognize your page image and apply the training file you just created.
- Click *No* if you want to return to the OmniPage screen without recognizing the image.

See “Deleting *.set, *.trn, *.ud, *.zcn, and *.zon Files” on page 236 for information on how to delete a training file.

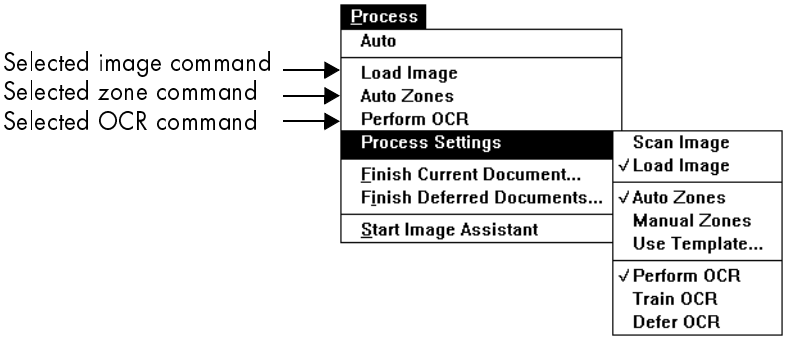
Process Settings

Choose *Process Settings* to access the image, zone, and OCR processing commands. You can also select these options in the Image, Zone, and OCR button drop-down lists.

The Process Settings cascading menu contains the following commands:

- *Scan Image* or *Load Image*
- *Auto Zones*, *Manual Zones*, or *Use Template...*
- *Perform OCR*, *Train OCR*, or *Defer OCR*

The currently selected Process Settings commands determine what image, zone, and OCR operations can be performed. OmniPage also uses the selected commands for automatic processing.



For more information on each *Process Settings* command, refer to its entry in this section starting on page 117.

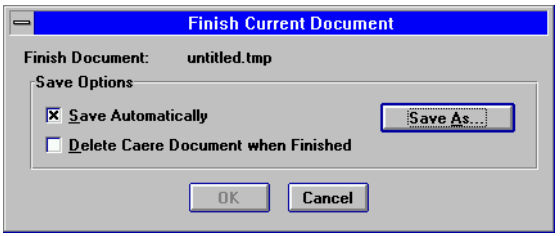
Finish Current Document

Choose *Finish Current Document...* to finish recognition of an open document.

OmniPage uses the currently selected Settings Panel options to finish processing your document.

Finishing the Current Document

1. Open the document you want to finish if it is not already open in OmniPage.
2. Choose *Finish Current Document...* in the Process menu.
   The Finish Current Document dialog box appears.



You can perform recognition and save the document later, or perform recognition and save the document automatically.

Perform OCR and Save Later

1. Deselect *Save Automatically* if it is selected.
2. Click *OK*.
   OmniPage recognizes your document and displays the text in the text window.
3. Choose *Save As...* in the File menu to save the recognized document when you want to do so.

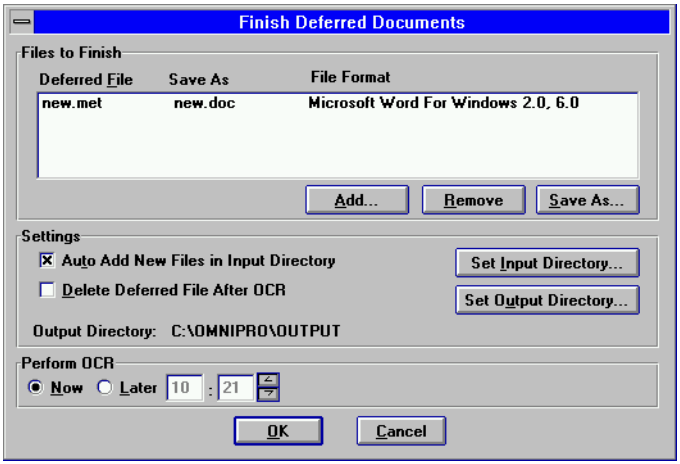
Perform OCR and Save Automatically

1. Select *Save Automatically* if it is not selected.
   This activates the other options in the dialog box.
2. Click *Save As...*.
   The Save As dialog box appears.
3. Select a file type for your saved file in the *Save Files as Type* drop-down list.
4. Type a name for your file in the *File Name* text box.
5. Select a location for your saved file.
6. Click *OK* to return to the Finish Current Document dialog box.
7. You can select *Delete Caere Document when Finished* to discard the Caere Document after recognition if desired.
   Your document will only be saved in the file format that you select. If the Caere Document is deleted, you cannot reopen or edit your recognized document in OmniPage.
8. Click *OK* to begin OCR.

OmniPage recognizes all pages and saves them as you specified. If a page does not have zones, OmniPage automatically creates zones using the selected Settings Panel Zones option.

Finish Deferred Documents

Choose *Finish Deferred Documents...* to finish OCR of your deferred documents. This opens the Finish Deferred Documents dialog box, described in the next section of this topic.



Choose *Defer OCR* in the Process menu or the OCR button drop-down list to defer page recognition during automatic processing. See “Defer OCR” on page 128 for more information.

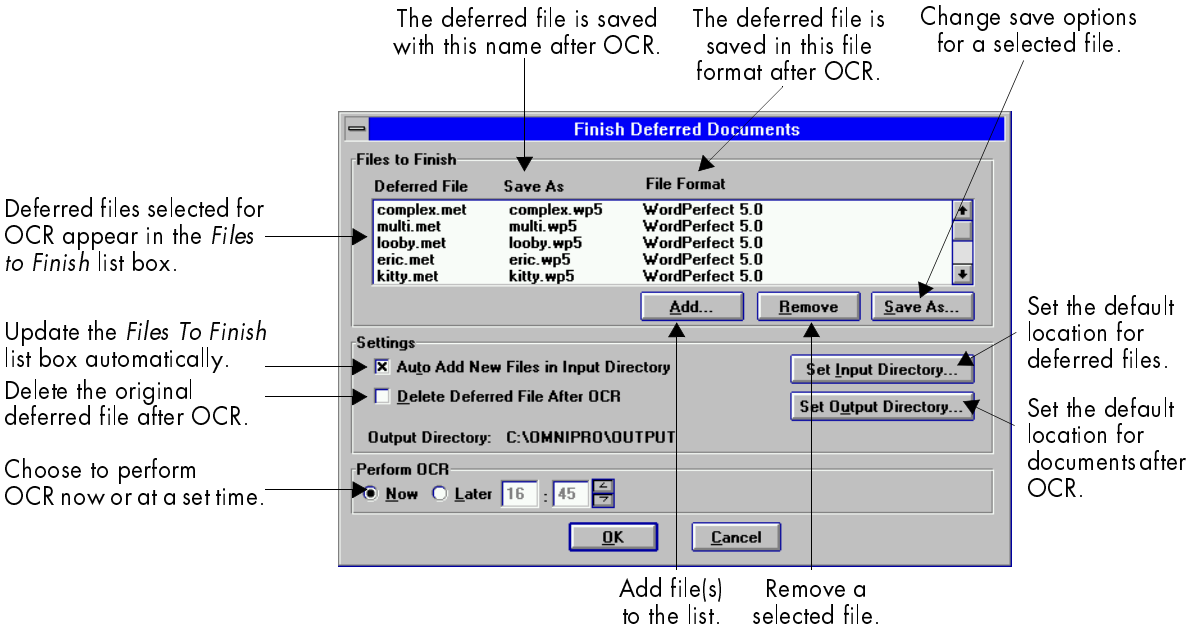
Deferred documents must be saved as Caere Documents or as image files in order to reopen and process them. You must have *Save Page Images in Caere Document* selected in the Preferences settings panel in order to save a scanned or loaded image for deferred processing.

OmniPage uses the currently selected Settings Panel options to process your deferred documents.

You can select as many as 100 documents at a time for deferred processing.

Finishing Deferred Documents

1 Choose *Finish Deferred Documents...* in the Process menu.
The Finish Deferred Documents dialog box appears.



You can now select one or more deferred files to recognize and set a variety of options to take place after recognition.

2 Click *Set Input Directory...*.
The Set Input Directory dialog box appears. It opens to the `omnipro\input` directory by default. This is the default directory in which OmniPage looks for your deferred files. You can select another location if you wish. The input directory cannot contain more than 100 documents for deferred processing.

3 Click *OK* to return to the Finish Deferred Documents dialog box.
All supported files in the input directory appear in the *Files to Finish* list box. See “Supported Input File Formats” on page 239 for a list of supported file formats. OmniPage assigns each file a format (based on the last-selected format to which you saved). It assigns each file a file name based on that format.

4 Click *Set Output Directory...*.
The Set Output Directory dialog box appears. It opens to the `omnipro\output` directory by default. This is the directory to

which files are saved automatically after OCR. You can select another location if you wish.

5 Click *OK* to return to the Finish Deferred Documents dialog box.

- Click *Add...* to add additional files to the *Files to Finish* list box.

  (You cannot add a file if there are already 100 files in the *Files to Finish* list box.) See "Supported Input File Formats" on page 239 for a list of file formats OmniPage supports.
- Click *Remove* to remove a selected file.
- Select a file and click *Save As...* to change the file type or location to which the file will be saved after recognition.
- Select any options you need under *Settings.*
- Select an OCR time, *Now* or *Later*.

  Set a time using the up / down arrows.

6 Click *OK*.

Click *Cancel* if you want to exit the dialog box without saving any changes and without performing OCR.

OmniPage opens and recognizes all documents listed in the *Files to Finish* list box. It saves each file with the name and file type you specified.

- Files are saved to the `output` directory after OCR unless you chose another location.
- The original deferred files are moved to the `output` directory after OCR.

  If you had selected *Delete Deferred File After OCR*, all deferred files in the input directory are deleted after OCR instead of being moved to the output directory.

OCR begins at the specified time if you selected *Later*. OmniPage must be running for OCR to take place when *Later* is selected.

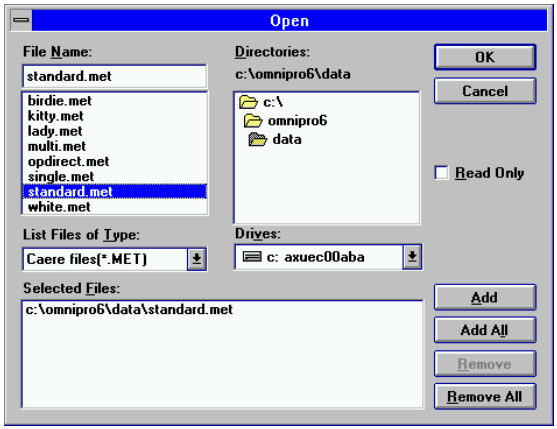
The following sections give detailed information on each component of the Finish Deferred Documents dialog box, listed alphabetically.

Add

Click *Add* to select deferred files for recognition. This could be a file outside your input directory, or a file added to the input directory when the *Auto Add New Files in Input Directory* option is deselected.

1 Click *Add...*.

2 The Open dialog box appears.



- Select a file type in the *List Files of Type* drop-down list to view specific image or deferred Caere Document (`*.met`) files.
- Select a file and click *Add* to add it to the *Selected Files* list box.
- Click *Add All* to add all files in a selected directory to the *Selected Files* list box.
- Select a file in the *Selected Files* list box and click *Remove* to remove it from the list.
- Click *Remove All* to remove all files from the *Selected Files* list box.

3 Click *OK* when you are done selecting files.

Click *Cancel* to close the dialog box without adding any files to the Files to Finish list box.

Selected files appear in the *Files to Finish* list box under the *Deferred File* header.

Auto Add New Files in Input Directory

Select this checkbox if you want OmniPage to update the *Files to Finish* list box automatically whenever new files are added to the input directory. Deselect this option if you do not want new files added automatically. You

can either click *Add...* to add one or more new files or *Set Input Directory...* to add all files in the input directory instead.

Delete Deferred File After OCR

Select this checkbox to delete all the original deferred files permanently from your hard drive after recognition. The original deferred files are moved to the output directory after recognition instead if this option is not selected.

If the Caere Document is deleted, you cannot reopen or edit your recognized document in OmniPage.

Files to Finish List Box

The *Files to Finish* list box lists the files in your input directory under the header *Deferred File*. Each file also has *Save As* and *File Format* options. This list is empty the first time you open the dialog box.

- You can click *Set Input Directory...* to choose a directory where OmniPage will look for your deferred documents.
- You can click *Set Output Directory...* to choose a directory to which the deferred documents will be saved after OCR.

Both these buttons are described later in this section.

Perform OCR

Select *Now* or *Later* in this section.

Select *Now* if you want OCR to begin as soon as you click *OK* in the Finish Deferred Documents dialog box.

Select *Later* and use the up / down arrows next to the clock numbers to set a time you want OCR to begin. The numbers are based on the 24-hour clock. 00:00 is 12 a.m. (midnight). 13:00 is 1 p.m., and so on. OmniPage must be launched for OCR to take place.

Remove

Select a file in the *Files to Finish* list box and click *Remove* to remove it from the list. This does not delete it from the directory unless the words *Job Failed* appear after the file name under *File Format*. Then a dialog box gives you the option of deleting the file as well as removing it. Click *Yes* to delete the file permanently or *No* to remove it from the *File to Finish* list box.

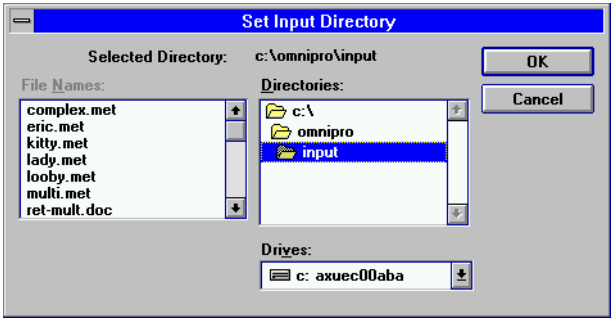
Save As

Select a file in the *Files To Finish* list box and click *Save As...* to open the Save As dialog box. You can determine how the selected file will be saved. Select a file type and change the file name if you wish. The new file name appears in the *Files to Finish* list box under the *Save As* header. The file format to which it will be saved appears in the *Files to Finish* list box under the *File Format* header.

See "Save Options" on page 92 for information on the way selected save options affect the length of the file name you can assign.

Set Input Directory

Click *Set Input Directory...* to set a default directory for your deferred files. This opens the Select Input Directory dialog box.



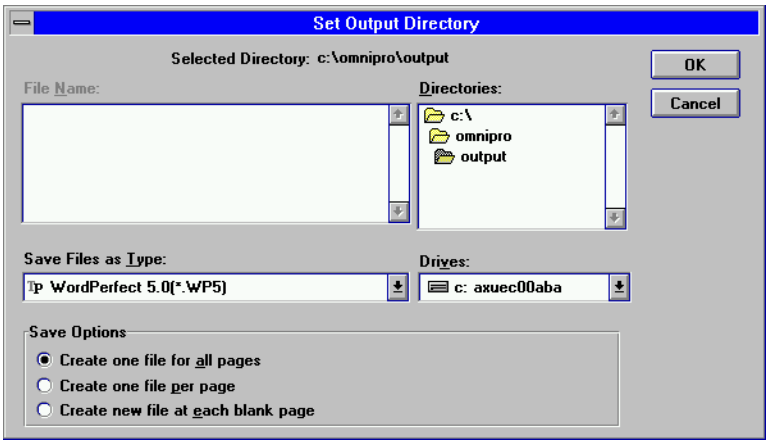
It opens to the `omnipro\input` directory by default. Click *OK* or select another directory from which OmniPage will open your deferred documents and click *OK*. Save your deferred Caere Documents and image format files (such as fax files) to this directory so that you can find and process them easily.

The files in the selected input directory will show up in the *Files To Finish* list box. Select the *Auto Add New Files in Input Directory* checkbox if you want the list to update automatically when new files are added to the input directory.

Set Output Directory

You can choose the default location for finished files.

1 Click *Set Output Directory...*.
The Set Output Directory dialog box appears.



The dialog box opens to the `omnipro\output` directory by default. You can select another directory to which OmniPage will save your finished documents after OCR if desired.

2 Select a file type in the *Save Files as Type* drop-down list.
The format you choose affects *all* the files in the *Files To Finish* list box.
You can select a file and click *Save As...* to change the file type or location to which the file will be saved after recognition as described previously in this section.

3 Select one of the *Save Options*.
The option you choose affects *all* the files in the *Files to Finish* list box.
See “Save Options” on page 92 for information.

OmniPage automatically truncates file names that are too long for the selected save option. If your file name can only have five characters (as with the *Create one file per page* option) but its name is `longer.doc`, for example, OmniPage would shorten it to `longe.doc`.

4 Click *OK* when you are done.

The Settings Menu

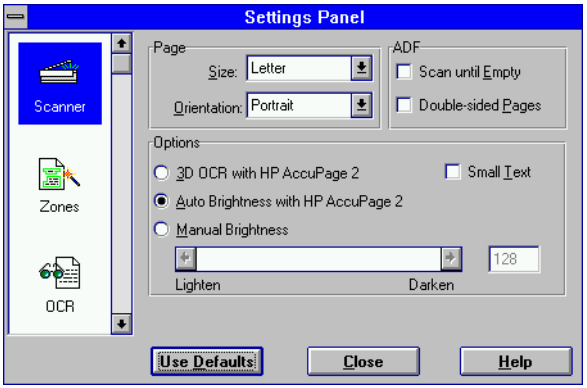
The Settings menu lets you modify system-wide settings. Settings menu commands include:

- Settings Panel
- Select Languages
- Register Applications
- Edit Training File
- Edit Zone Contents File
- Edit User Dictionary

Settings Panel

Choose *Settings Panel...* to open the Settings Panel. This command is also available as a button in the toolbar.

The Settings Panel is the central location for document processing settings.



There are seven panels in the Settings Panel dialog box: Scanner, Zones, OCR, Fonts, Spelling, Direct Input, and Preferences. Click each icon on the left to display its settings.

The Settings Panel changes to reflect the options of the icon that you click. Click *Help* to open the online Help program to information on the currently selected Settings Panel options.

For detailed information about these options, see Chapter 4, The Settings Panel.

Select Languages

Choose *Select Languages...* to select one or more language character sets for text recognition.

OmniPage can recognize additional characters (such as circumflexes, umlauts, etc.) unique to a particular language; thirteen languages are available. You may select more than one language at a time, but for faster recognition, use only the minimum number of languages that are necessary.

Select the main dictionary that matches your language selection. Deselect *Use Language Analyst* in the OCR settings panel and *3D OCR* in the Scanner settings panel if you do not have the appropriate dictionary. Select only one language if you use the Language Analyst or *3D* OCR.

Selecting a Language

1. Click once on a language to select it.
2. Select as many languages as you need.
   Selected languages are highlighted.
   To deselect a language, click it again.
3. Click *OK*.
   Click *Cancel* to exit without setting the selected language(s).

Register Applications

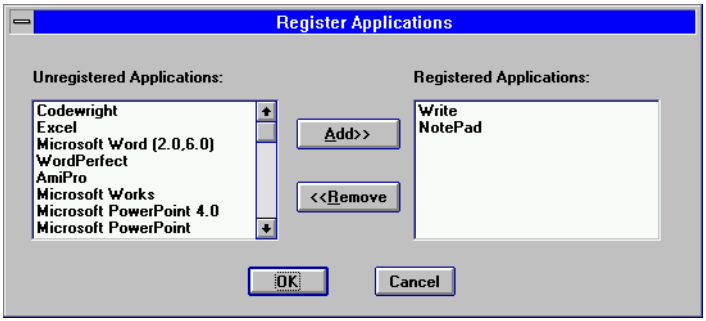
Choose *Register Applications...* to select which of your applications will support Direct Input or to remove an application from the *Registered Applications* list.

Direct Input lets you launch OmniPage from a supported application and place recognized text directly in your open application. (Do not confuse this command with the Register menu that appears if you have not registered OmniPage.) Two Windows applications, Write and Notepad, are pre-registered.

See Chapter 5, Direct Input, for detailed information.

To register an application:

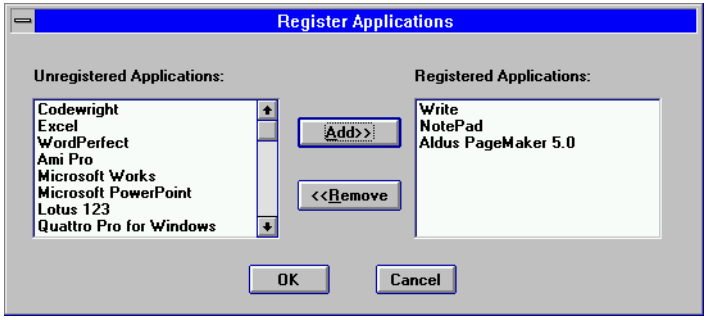
1 Choose *Register Applications* in the Settings menu.
The Register Applications dialog box appears.



Applications that support Direct Input appear in the *Unregistered Applications* list box. Applications appear in the left list box whether or not they are installed on your system. An application must be installed, however, to use it to initiate Direct Input.

2 Click any application on the left to select it.

3 Click *Add>>*.
The application moves into the *Registered Applications* list box.



4 Move as many applications as you would like to register.

5 Click *OK* when you are done.

OmniPage immediately modifies the File menu of a registered application to contain the *Direct Input...* command when that application is open.

You would choose *Direct Input...* in your application's File menu to launch OmniPage in Direct Input mode.

Edit Training File

Choose *Edit Training File...* to edit an existing character training file.

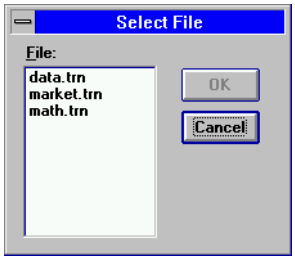
A character training file (`*.trn`) is a set of pre-recognized text characters that OmniPage compares to the characters in the page image during recognition. Training files assist OmniPage during text recognition and allow better recognition accuracy of special characters.

See "Train OCR" on page 129 for detailed information on how to create a training file.

See "Deleting *.set, *.trn, *.ud, *.zcn, and *.zon Files" on page 236 for information on how to delete a training file.

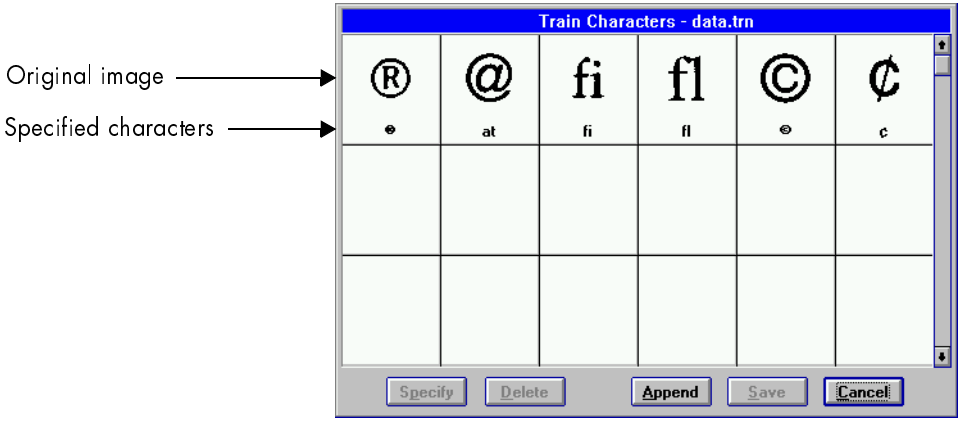
Editing a Training File

1 Choose *Edit Training File...* in the Settings menu.
The Select File dialog box appears.



2 Select a training file.

3 Click *OK*.
The Train Character dialog box appears.



The Train Character dialog box shows the existing characters in the training file, including the original images and the associated characters. You can edit characters, delete them, or append this file to another one.

- Select a character in the dialog box and click *Specify* to change how it will be recognized by OmniPage.

  See "Specify" on page 130 for detailed information.

- Select a character and click *Delete* to discard it from the training file.

4 Do one of the following:

- Click *Save* to save the edited training file and return to the OmniPage screen.
- Click *Append* to add the trained characters to another file.

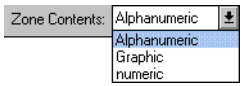
  See "Append" on page 131 for detailed information.

- Click *Cancel* to exit without saving the edits to the training file.

Edit Zone Contents File

Choose *Edit Zone Contents File...* to create a new zone contents file or edit an existing file.

A zone contents file (`*.zcn`) lets you identify the specific characters that OmniPage looks for within specified zones during recognition. OmniPage is shipped with numeric, graphic, and alphanumeric zone contents files. All zone contents files appear in the *Zone Contents* drop-down list in the zone window.

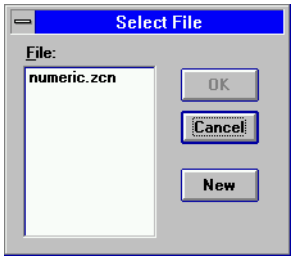


- The Numeric file is editable. Characters can be added to or deleted from this file.
- Any zone recognized as a graphic can be exported as a separate graphic file.
- You can create additional zone contents files with the characters you require.

See "Deleting *.set, *.trn, *.ud, *.zcn, and *.zon Files" on page 236 for information on how to delete a zone contents file.

Creating a New Zone Contents File

1 Choose *Edit Zone Contents File...* in the Settings menu.
The Select File dialog box appears.



2 Click *New*.
The Edit Zone Content File dialog box appears.

The *Zone Contents* edit box contains the 94-character (typical keyboard) ASCII character set. More characters are in the *Extended ANSI* list box to the right.

- Click in the *Zone Contents* edit box and delete or add characters as needed.

If you do not click first to deselect the edit box, any character you type will replace the entire zone contents set. Click *Reset* if this happens.

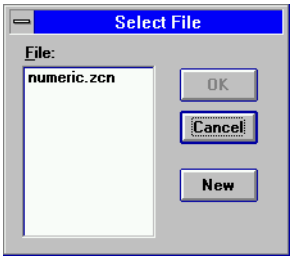
- Double-click any character in the *Extended ANSI* list box to add it to the *Zone Contents* text box.

3 Click *Save* when you are done.
The File to Save dialog box appears.

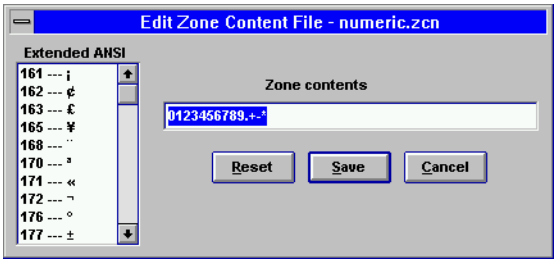
4 Type a name in the edit box and click *OK*.

Editing a Zone Contents File

1 Choose *Edit Zone Contents File...* in the Settings menu.
The Select File dialog box appears.



2 Select a file to edit and click *OK*.
The Edit Zone Content File dialog box appears.



The *Zone Contents* edit box contains characters you specified previously. More characters are in the *Extended ANSI* list box to the right.

- Click in the *Zone Contents* edit box and delete or add characters as needed.

If you do not click first to deselect the edit box, any character you type will replace the entire zone contents set. Click *Reset* if this happens.

- Double-click any character in the *Extended ANSI* list box to add it to the *Zone Contents* text box.

3 Click *Save* when you are done.

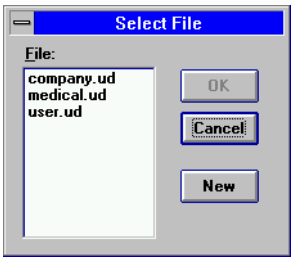
Edit User Dictionary

Choose *Edit User Dictionary...* to create a new user dictionary (`*.ud`) or edit an existing one. The `user.ud` is the default user dictionary supplied with OmniPage.

Creating or Editing a User Dictionary

1 Choose *Edit User Dictionary...* in the Settings menu.

The Select File dialog box appears.



The dialog box lists all user dictionary files in the `data` directory.

- To edit an existing dictionary, select a file and click *OK*.
- To create a new dictionary, click *New*.

  Enter a name in the dialog box that appears and click OK.

Whether you are creating a new dictionary or editing an existing one, the Edit User Dictionary dialog box appears.

Words in the dictionary appear in the list box at the left. No words appear in a new dictionary.

2 Use the buttons in the Edit User Dictionary dialog box to create or edit your dictionary:

- Click *Add* to add the word that you type in the *User word* edit box to your dictionary. The word will appear in the list box.
- Click *Delete* to delete a selected word from the dictionary.

- Click *Delete All* to delete all words from the dictionary.
- Click *Import...* to add words from another application to your user dictionary. For example, you may want to add technical terms from a particular file.

  The Import Text File dialog box prompts you to enter the file name and directory of the file you want to import. An imported text file can be any document or word list in ASCII format. Most word processors can convert a file into ASCII format; see your program's documentation.

  OmniPage will go through the selected text file, discard words already in the main or other user dictionaries, and add the remaining words to your current user dictionary.
- Click *Export...* to save your user dictionary as a text file. The Export To Text File dialog box prompts you to enter a file name and destination for your file.

  The default location is `omnipro\data`. This is the only place OmniPage looks for these files.
- Click *Save As...* to save an edited dictionary with a new name. User dictionaries are automatically saved with a `*.ud` file extension.

3 Click *Done* to save edits to your dictionary and then exit.

See "Deleting *.set, *.trn, *.ud, *.zcn, and *.zon Files" on page 236 for information on how to delete a user dictionary file.

The Register Menu

There are no commands in this menu. It *only* appears in the menu bar if you have not registered your copy of OmniPage. Do not confuse this with the *Register Applications...* command in the Settings menu.

1 Click the Register menu to open the OmniPage Product Registration dialog box.

You can use OmniPage for 25 sessions without registering. After the 25th time, the Registration dialog box appears on launch but if you click *Cancel*, the program exits.

Registering your copy of OmniPage entitles you to technical support, notification of special offers and upgrades, and the lowest price offered on the next OmniPage upgrade.

2 Click the *Call* drop-down list to find the number you should call from your country.

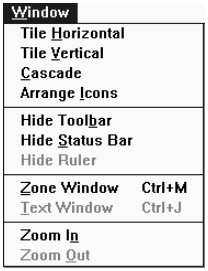
3 Call the number and ask for a registration number.

You will be asked to provide some information and you will be assigned a registration number.

4 Enter the number in the *Registration Number* text box and click *OK*.

You are now a registered user of OmniPage. The Registration menu disappears from the menu bar.

The Window Menu



The Window menu provides options for viewing the OmniPage screen and your document. Window menu commands include:

- Tile Horizontal
- Tile Vertical
- Cascade
- Arrange Icons
- Hide/Show Toolbar
- Hide/Show Status Bar
- Hide/Show Ruler
- Zone Window
- Text Window
- Zoom In
- Zoom Out

Tile Horizontal

Choose *Tile Horizontal* to resize the open zone and text windows so they fit in the window area horizontally. To switch windows, click the window that you want to activate.

Tile Vertical

Choose *Tile Vertical* to resize the open zone and text windows so they fit in the window area vertically. To switch windows, click the window that you want to activate.

Cascade

Choose *Cascade* to arrange the open zone and text windows one on top of the other with title bars showing. To switch windows, click the title bar of the window you want to activate.

Arrange Icons

Choose *Arrange Icons* to organize minimized window icons at the bottom of the screen. Click the Minimize button in the upper-right corner of the window to iconize the open zone and text windows.

Hide/Show Toolbar

Choose *Hide Toolbar* to hide the toolbar. Choose *Show Toolbar* to view the toolbar again.

Hide/Show Status Bar

Choose *Hide Status Bar* to hide the status bar located at the bottom of the window. Choose *Show Status Bar* to view the status bar again.

Hide/Show Ruler

Choose *Hide Ruler* to hide the text window ruler. Choose *Show Ruler* to view the ruler again.

Zone Window

Choose *Zone Window* to bring the zone window into view.

Text Window

Choose *Text Window* to bring the text window into view.

Zoom In

Choose *Zoom In* to enlarge an area of an image in the zone window for a close-up view. You can also click the Zoom tool in the zone window and click anywhere in the image to zoom in.

When an image is opened, it is fit to the zone window. You can zoom in three more levels.

Zoom Out

Choose *Zoom Out* to decrease an enlarged view of an image in the zone window. You can also click the Zoom tool in the zone window and click anywhere in the image with your *right* mouse button to zoom out.

The Help Menu

The Help menu provides access to the OmniPage online Help program and information about OmniPage. Help menu commands include:

- Contents
- Procedures
- Using Help
- About

Contents

Choose *Contents* for a list of the topics available in the OmniPage Help program. The Help program conforms to the Windows Help standard.

Procedures

Choose *Procedures* for a Help listing of procedures for different OmniPage tasks.

Using Help

Choose *Using Help* for instructions on using the Help program.

About

Choose *About...* to see information about the current OmniPage version, any copyrights in effect, the program's licensee, company name, and serial number.

Chapter 4

The Settings Panel

This chapter explains the way the settings you select in the Settings Panel affect OmniPage document processing.

The following topics are included:

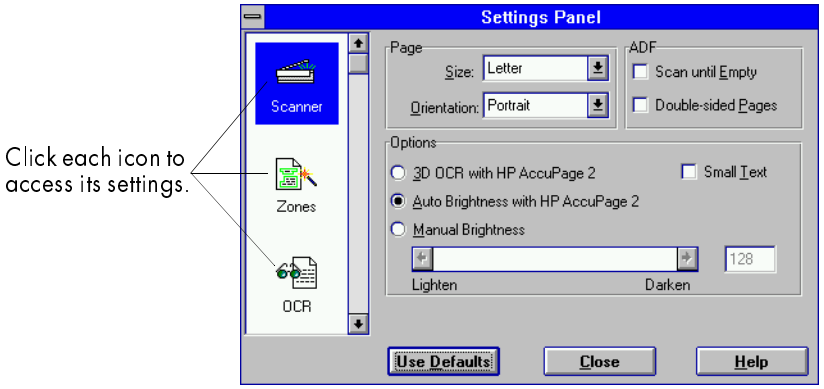
- Settings Panel Overview
- Scanner Options
- Zones Options
- OCR Options
- Fonts Options
- Spelling Options
- Direct Input Options
- Preferences Options

The options in each settings panel are described from left to right, top to bottom.

Some Settings Panel options are unavailable in Direct Input mode. Refer to Chapter 5, Direct Input, for detailed information.

Settings Panel Overview

To open the Settings Panel, choose *Settings Panel...* in the Settings menu or click the Settings Panel button in the toolbar.



Each one of the seven icons in the scroll box on the left side of the Settings Panel represents a different set of options.

Click the Scanner icon to select options that control the way your scanner scans a page and the way an image file is loaded.

Click the Zones icon to select the zoning option that determines how text is organized during recognition.

Click the OCR icon to select input options that assist OmniPage during recognition and output options to set formatting and page layout.

Click the Fonts icon to select font format options for retaining or ignoring the original font styles.

Click the Spelling icon to select spell-checking options and dictionaries for recognition languages.

Click the Direct Input icon to select processing and output options for scanning text directly into an open application.

Click the Preferences icon to select options that customize general OmniPage operations.

The Settings Panel changes to reflect the options of the icon that you click. You can select options and then click *Close* or leave the Settings panel open as a floating window. OmniPage retains the last-selected options until you select new ones.

Settings Panel Buttons

There are three buttons in the Settings Panel: *Use Defaults*, *Close*, and *Help*.

- Use Defaults

  Click *Use Defaults* to return options for *all* settings panels to their default state. A dialog box prompts you to confirm your selection if you have changed any setting from its default.

- Close

  Click *Close* to close the Settings Panel. Settings Panel options take effect once selected even if the Settings Panel is still open. (*Save Settings on Quit* is selected by default in the Preferences settings panel.) If you exit OmniPage and leave the Settings Panel open, however, changed settings are not saved.

  For example, suppose *Multiple Columns* is selected when you open the Settings Panel. You open the Settings Panel, select *One Zone*, and exit OmniPage without closing the Settings Panel. The next time you launch OmniPage, *Multiple Columns* is selected again.

- Help

  Click *Help* to open context-sensitive help. If you click *Help* while the OCR settings panel is open, for example, the online Help program opens to the topic titled “OCR Options.”

Selecting Settings Panel Options

There are three ways to select Settings Panel options:

- Manual selection

  Click each Settings Panel icon and select options manually. You can change your selections at any time.

- *Use Defaults* button

  Click the *Use Defaults* button in the Settings Panel to reset options for *all* settings panels to the default values. A dialog box appears to confirm your choice.

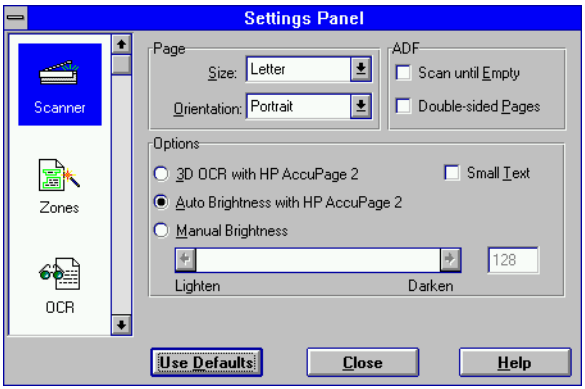
- *Load Settings* command

  You can save Settings Panel selections and a language character set to a settings file. Choose *Load Settings...* in the File menu to load the file. See “Save Settings” on page 98 and “Load Settings” on page 99 for more information. See “Deleting *.set, *.trn, *.ud, *.zcn, and *.zon Files” on page 236 for information on deleting a settings file.

Scanner Options

Click the Scanner icon in the Settings Panel to select options that control the way your scanner scans a page.

Use your right mouse button to click the Image button in the toolbar and automatically open the Settings Panel to Scanner options.



Page

Select *Page* options to describe the page size and orientation of the document you want to recognize.

Size

The *Size* drop-down list box lets you select the dimensions of the pages you are scanning.

- Select *Letter* for 8.5" by 11" size pages.
- Select *Legal* for 8.5" by 14" size pages.
- Select *A4* for 21 cm by 29.7 cm European-size pages.

Orientation

The *Orientation* drop-down list box lets you select the orientation of the pages you scan or images you load. Be sure to position the pages correctly in the scanner when scanning.

Select *Portrait* for a vertically oriented page.

Select *Landscape* for a horizontally oriented page.

Select *Flipped* to automatically rotate a portrait page 180 degrees during the scan.

Select *Flipscape* to automatically rotate a landscape page 180 degrees during the scan.

Flipped and *Flipscape* options are useful if you are scanning pages in a book and need to turn the book upside down or sideways for certain pages.

ADF

You can select *Scan until Empty* and *Double-sided Pages* for automatic processing if you are using a scanner with an automatic document feeder (ADF).

Scan until Empty

Select this to scan every page in an ADF when OmniPage performs automatic processing.

When you place multiple pages in an ADF and click the *AUTO* button, the first page is scanned and processed and then the next page. This process continues until the ADF is empty. OmniPage uses the same settings and processing commands for each page.

You must click *AUTO* to process all pages in a stack; if you click the Image button instead, only the first page is scanned. No matter which button you click, OmniPage scans only the first page in an ADF if you do not select *Scan Until Empty*.

Double-sided Pages

Select this when OmniPage performs automatic processing to scan pages with print on both sides. OmniPage processes all pages in the ADF and then prompts you to turn the entire stack over to process the reverse sides.

For example, if you have three double-sided pages numbered 1 through 6 (1 is on the front, 2 is on the back, and so on), OmniPage first processes pages 1, 3, and 5 and then prompts you to turn the stack over in the ADF.

It then processes pages 6, 4, and 2. The resulting file consists of pages 1, 2, 3, 4, 5, and 6 in the correct order.

You can divide a large batch of pages into several stacks for processing. OmniPage processes one side of the first stack and then the reverse side. It repeats this procedure with the second stack of pages, appends it to the first stack of pages, and so on for any new stacks. If you want to save each stack as a separate file, insert blank pages as separators. See "Save Options" on page 92 for information on how the pages are saved.

Do not select *Double-sided Pages* if you use a flatbed scanner without an ADF. Place the pages in the scanner one by one in the order that you want them to be scanned.

Options

Select a brightness setting for scanning your page. Use these options to account for variations in paper and print quality in much the same way you would adjust brightness on a copier. Depending on the quality of your page, the option you choose can greatly affect recognition accuracy.

You can select:

- *3D OCR with AnyPage/HP AccuPage 2*
- *Auto Brightness with AnyPage/HP AccuPage 2*
- *Manual Brightness*.

Auto Brightness with HP AccuPage 2 is available with newer HP scanners.

Auto Brightness with AnyPage is available with all other supported grayscale scanners.

Manual Brightness is the only option available to users who have black-and-white scanners.

3D R with AnyPage/HP AccuPage 2

Select this to combine *3D* OCR and AnyPage / HP AccuPage 2 technologies for the best scanned image and highest recognition accuracy possible. This option is only available with supported grayscale scanners.

3D OCR technology uses the grayscale information on a page during recognition to view characters clearly and completely. AnyPage / HP AccuPage 2 technology automatically adjusts an image to get the optimum brightness level for each area of text and graphics on a page.

When to Use

Use *3D OCR and AnyPage/HP AccuPage 2* whenever you want the best possible recognition results. This setting is especially useful when you scan:

- poor-quality pages
- pages with very small type
- pages with text on colored or shaded backgrounds

When to Use Another Setting

This setting is slower than the other two settings. Select *Manual Brightness* for the fastest results if you scan high-quality documents with crisp black text on a white background.

3D OCR adds 150 to 250K per page to the size of the image file. Use the *Auto Brightness with AnyPage/HP AccuPage 2* or *Manual Brightness* if you need to save disk space.

Auto Brightness with AnyPage/HP AccuPage 2

Select this to use AnyPage / HP AccuPage 2 technology to get high-quality scanned images and high recognition accuracy. This option is only available with supported grayscale scanners.

When to Use

AnyPage / HP AccuPage 2 technology automatically adjusts images to get the optimum brightness level for each area of text and graphics on a page.

This setting works well for most pages but is especially useful for text on colored or shaded backgrounds.

When to Use Another Setting

This setting is slower than a manual setting. If you scan high-quality documents with crisp text on a white background, select *Manual Brightness* for the fastest results.

Manual Brightness

Select this to manually adjust (lighten or darken) the brightness setting for scanning a page. The setting you choose is applied to the entire page area.

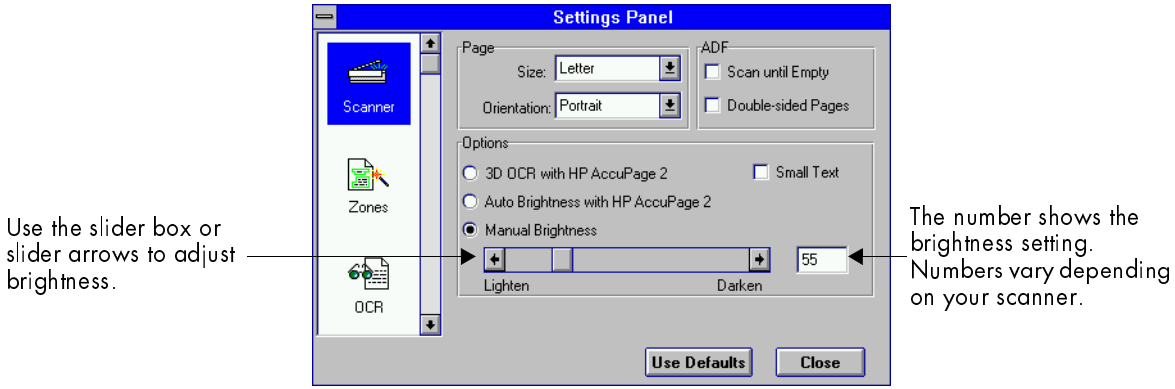
When to Use

Manual Brightness is the fastest setting if you scan high-quality documents with crisp text on a white background.

How to Use

To adjust brightness, select the slider box, hold the mouse button down, and move the square to lighten or darken the setting. Or, click the left or

right arrow on the slide. The number of settings available depends on the scanner you use.



- Use a setting in the middle to scan high-quality documents with crisp text on a white background.
- Use a darker setting for a page that has thin, broken characters.
- Use a lighter setting for a page that has thick, run-together characters.

The number in the edit box to the right of the slide shows the brightness level you selected. Use this number as a reference for future use.

Watch the Character window during OCR and look for clear, legible text. See “Scanning and the Brightness Setting” on page 53 for examples.

When to Use Another Setting

Recognition accuracy is highest using AnyPage / HP AccuPage 2 and *3D* OCR technologies. These two options are only available if you have a supported grayscale scanner chosen as the default scanner.

Small Text

Select this option if your scanner supports HP AccuPage 2 and the text in your page is between four and seven points. This option does not appear for other scanners. The other settings under *Options* are unavailable when *Small Text* is selected.

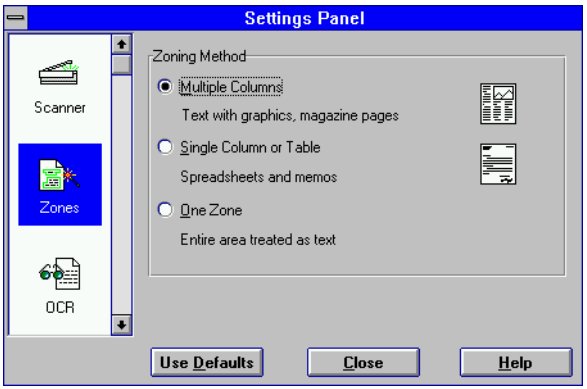
When to Use

Select *Small Text* to increase recognition accuracy. This setting increases processing time slightly.

Zones Options

Click the Zones icon in the Settings Panel to select the zoning method that determines the flow of text during recognition.

Use your right mouse button to click the Zone button in the toolbar when it is active and automatically open the Settings Panel to Zones options.



You can create zones manually, automatically, or with a template, but OmniPage still uses the selected zoning method in the Settings Panel to determine text flow within each zone. OmniPage also uses the selected zoning method to draw and order zones on the page image when you choose the *Auto Zones* processing command.

Multiple Columns

Select *Multiple Columns* if you want OmniPage to discern the column layout, determine the order of text, and distinguish graphics from text. This works well with most types of documents and is especially useful for:

- Multiple columns
- newspaper and newsletter pages
- magazine pages

How it Works

OmniPage looks for regular vertical separations of text to define columns and then recognizes column-wide text zones. It starts at the top of the first column, moves to the bottom, then continues to the top of the next column.

Select *Retain Graphics* in the OCR settings panel when you select *Multiple Columns* to preserve your graphics. Otherwise, graphics are discarded.

Multiple Columns and OCR Settings

Select *Multiple Columns* along with *True Page* in the OCR settings panel if you want to reproduce page layout as closely as possible and retain font and paragraph formatting. These options also let you save recognized text in a frame-based format. See Chapter 6, Using True Page, for information.

Select *Multiple Columns* along with *Retain Font and Paragraph Formatting* in the OCR settings panel if you want recognized text displayed in one column with any retained graphics at the end. OmniPage retains font and paragraph formatting but does not attempt to reproduce page layout.

Select *Multiple Columns* along with *Ignore All Formatting* in the OCR settings panel if you want recognized text displayed in one column with any retained graphics at the end. OmniPage ignores font and paragraph formatting and does not attempt to reproduce page layout.

Single Column or Table

Select *Single Column or Table* if you want OmniPage to treat the entire page area as one column. This works well with documents such as:

- spreadsheets
- tables and charts
- financial forms
- memos and letters

How it Works

OmniPage starts at the top of the page and moves to the bottom, outlining page-wide text zones. If OmniPage detects five or more spaces between columns, it assumes the page is in a spreadsheet, or tabular, format and inserts tabs as delimiters between the columns to preserve the format.

Retaining Graphics

You must draw zones manually, identify graphics with the *Graphic* zone contents drop-down list, and select *Retain Graphics* in the OCR settings panel to retain graphics with this option. They are discarded otherwise.

Retain Graphics is disabled if *Save Page Images in Caere Document* in the Preferences settings panel is not selected before you scan or load an image.

Single Column or Table and OCR Settings

Select *Single Column or Table* along with *True Page* in the OCR settings panel to retain font and paragraph formatting, and separate distinct vertical spaces of five or more with tabs and distinct horizontal spaces with line returns. Unlike documents processed with the *Multiple Columns* with *True Page* options, those processed with the *Single Column or Table* and *True Page* options are exported without frames. See Chapter 6, Using True Page, for information.

Select *Single Column or Table* along with *Retain Font and Paragraph Formatting* in the OCR settings panel to have OmniPage separate distinct vertical spaces of five or more with tabs but ignore extra horizontal spacing. Font and paragraph formatting is retained.

Select *Single Column or Table* along with *Ignore All Formatting* in the OCR settings panel to have OmniPage separate distinct vertical spaces of five or more with tabs and ignore horizontal spacing and font and paragraph formatting.

One Zone

Select *One Zone* if you want OmniPage to recognize the entire page area as a single text zone. This option does not retain graphics even if they are defined.

How it Works

OmniPage does not reproduce page layout or distinguish graphics from text. It draws a single zone around the entire page when you use auto zoning and tries to recognize everything it sees on the page as text. *One Zone* is the fastest option to use if you recognize manually drawn, text-only zones. It is useful for documents that consist entirely of text such as:

- pleading pages
- telephone book pages (draw zones manually around each column)

One Zone and OCR Settings

Select *One Zone* along with *True Page* in the OCR settings panel if you want to retain font and paragraph formatting, and separate distinct vertical spaces of five or more with tabs and distinct horizontal spaces with line returns. Unlike documents processed with the *Multiple Columns* and *True Page* options, documents processed with the *One Zone* and *True Page* options are exported without frames. See Chapter 6, True Page, for more information.

Select *One Zone* along with *Retain Font and Paragraph Formatting* in the OCR settings panel to retain font and paragraph formatting, and separate

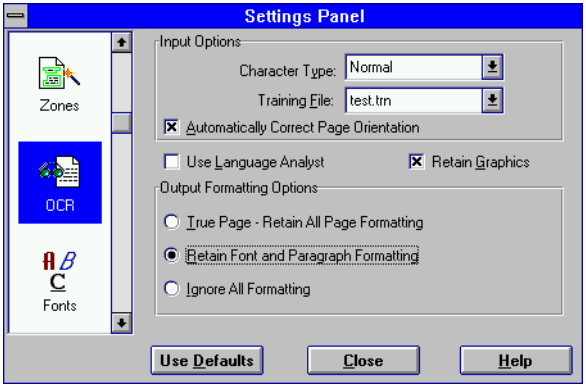
distinct vertical spaces of five or more with tabs but ignore extra horizontal spacing.

Select *One Zone* along with *Ignore All Formatting* in the OCR settings panel to ignore font and paragraph formatting but separate distinct vertical spaces of five or more with tabs.

OCR Options

Click the OCR icon in the Settings Panel to select input and output options that assist OmniPage during recognition and determine the format of the recognized document.

Use your right mouse button to click the OCR button in the toolbar when it is active and automatically open the Settings Panel to OCR options.



Input Options

Input options determine the way OmniPage looks at text elements during recognition. You can specify a character type, select a training file, and have the page orientation automatically corrected.

Character Type

The *Character Type* drop-down list box lets you identify the printed text characteristics in your document.

- Select *Automatic* to have OmniPage automatically distinguish between conventional and dot-matrix printed text characters in the image you are recognizing.

- Select *Normal* if the image you are recognizing has conventionally printed text characters.
- Select *Dot Matrix* if the image you are recognizing has characters printed in draft mode by a 9-pin dot-matrix printer. Do not select *Dot Matrix* for pages printed in near-letter-quality mode or printed by a 24-pin dot-matrix printer.
- Select *OCR-A* if all the characters in the image you are recognizing are printed in OCR-A font. OCR-A is a special font used for items such as part numbers and utility bills.

If your document contains a mixture of OCR-A and a conventional font, select *Normal* or *Automatic* for faster recognition.

Training File

Click the *Training File* drop-down list to select a character training file (`*.trn`). A training file assists OmniPage with text recognition of special characters. Any training files that you create appear in this list.

A character training file is a set of pre-recognized text characters that OmniPage compares with the characters in the page image during recognition. For more information on creating a character training file, please see “Train OCR” on page 129.

Automatically Correct Page Orientation

Select this to correct an improperly oriented image by 90, 180, or 270 degrees during text recognition. For example, if a page was accidentally scanned upside-down, OmniPage will try to rotate it 180 degrees during recognition so that it is properly oriented in the text window.

Use Language Analyst

Select *Use Language Analyst* so that OmniPage automatically performs word and character analysis during recognition to check spelling and replace unknown words with the words most likely to be correct.

How it Works

The Language Analyst uses information about language context and usage rules to evaluate words, find likely errors, and replace words. Replacement words appear in blue in the recognized document.

Be sure to select the appropriate main and user dictionaries in the Spelling settings panel when you use the Language Analyst. Select the appropriate language(s) in the Select Languages dialog box (choose *Select Languages...* in the Settings menu).

The Language Analyst shuts itself off automatically when it detects that the dictionary information is not improving recognition results. For example, if the main dictionary does not match the primary language of your document, language analysis stops.

If your original is very clean with crisp text, making character recognition easy, you may want to deselect the Language Analyst to increase recognition speed.

Inappropriate Word Replacement

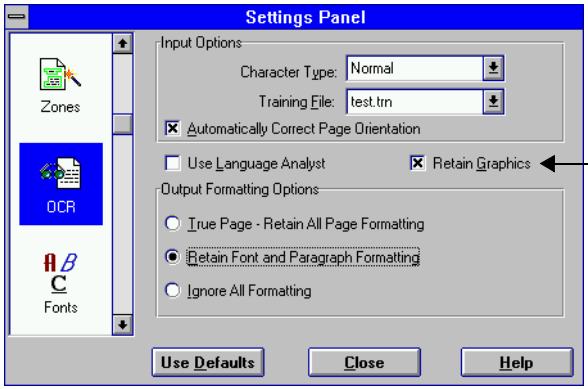
If any words in your document such as company-specific terms are replaced inappropriately during recognition, you can:

- Make sure *Ignore Acronyms*, *Ignore Abbreviations*, and *Ignore Proper Nouns* are selected in the Spelling settings panel so that these types of words will not be replaced. Then, re-recognize the document.
- Create you own user dictionary for special terms and select it as the user dictionary in the Spelling settings panel. Then, re-recognize the document.

  See “Edit User Dictionary” on page 149 for more information on user dictionaries.
- Deselect *Use Language Analyst* and re-recognize the document. Use the *Check Recognition* command in the Edit menu to check for spelling errors and unknown words.

Retain Graphics

Select *Retain Graphics* if you want OmniPage to retain original graphics such as photographs or drawings in the recognized document.



How to Retain Graphics

In addition to selecting *Retain Graphics* in the OCR settings panel, you must:

- Select *Save Page Images in Caere Document* in the Preferences settings panel before you scan or load an image.

  Retain Graphics is disabled if *Save Page Images in Caere Document* is not selected.

- Select *Multiple Columns* in the Zones settings panel so OmniPage automatically distinguishes graphics from text.

  If you select *Single Column or Table* or *One Zone* in the Zones settings panel, you must create zones manually and identify graphics with the *Graphic* zone contents file.

Select the *True Page – Retain All Page Formatting* OCR output option to keep graphics in the same position in the recognized page that they were in the original page. Retained graphics are placed at the bottom of the recognized page otherwise.

Output Options

Output options determine the way text, graphics, and formatting will appear in the recognized document. You can select:

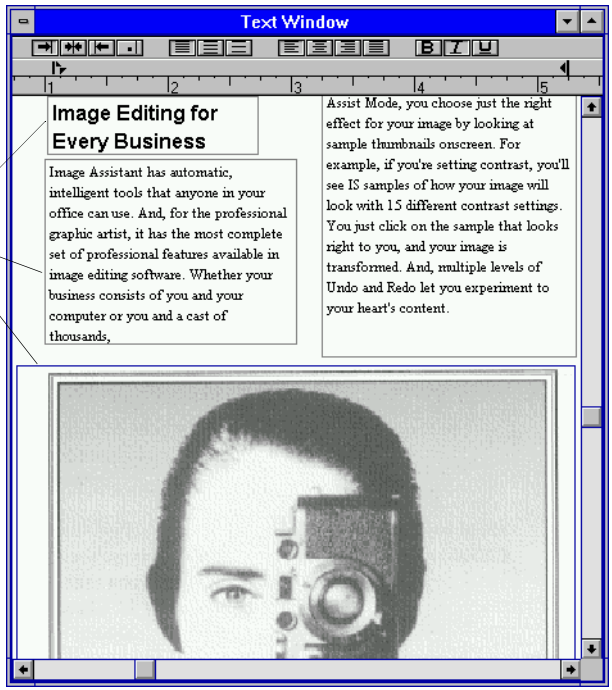
- *True Page – Retain All Page Formatting*
- *Retain Font and Paragraph Formatting*
- *Ignore All Formatting*

Files saved in ASCII or ANSI format do not retain any formatting other than spaces and carriage returns.

See “Zones Options” on page 163 for information on how Zones and OCR settings panel options work together during OCR.

True Page – Retain All Page Formatting

Select *True Page - Retain All Page Formatting* as the OCR output option if you want the recognized document to match the original page layout as closely as possible.



The True Page setting preserves the text window page layout in frames when you also select the *Multiple Columns* zoning method.

Select True Page formatting only if you want to preserve the original page layout. If you plan to do extensive editing, such changing page layout, select *Retain Font and Paragraph Formatting* or *Ignore All Formatting.*

What True Page Reproduces

True Page attempts to reproduce the following during page recognition:

- Relative text column positioning
- Relative graphic positioning (select *Retain Graphics* in the OCR settings panel)
- Margins
- Tabs
- Line Spacing
- Paragraph groupings
- Indentation
- Text alignment (justified, centered, or left-reading text)
- Blank vertical space
- Font styles (select specific fonts in the Fonts settings panel)
- Font sizes
- Character styles (boldface, italics, underline)

See Chapter 6, True Page, for more information.

Exporting Frames From a Recognized Document

True Page preserves layout by creating *frames* around areas of text and graphics if *Multiple Columns* is selected in the Zones settings panel. The frames are exported intact when you save a document in an appropriate file format and open it in an application that supports frame-based layouts.

File formats that support frame-based output from True Page have the letters *TP* in front of their name in the Save As dialog box.

If you want to preserve as much formatting as possible in a multiple-column document, but do not want exported text to have frames, choose *Retain Font and Paragraph Formatting* in the OCR settings panel along with *Multiple Columns* in the Zones setting panel.

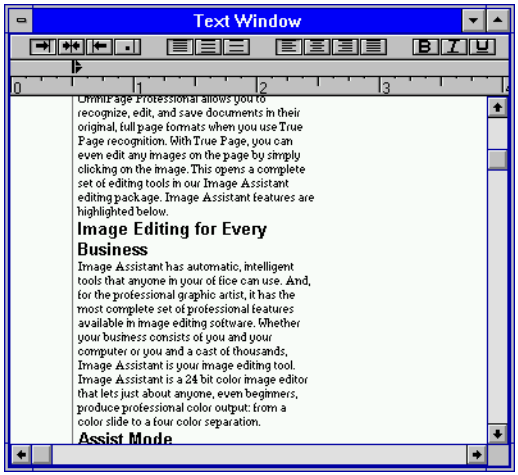
Exporting no Frames From a Recognized Document

True Page retains page layout but does not export recognized text with frames if you chose either *Single Column or Table* or *One Zone* in the Zones settings panel. The layout will vary depending on the format to which you save the document text.

Retain Font and Paragraph Formatting

Select this to retain font characteristics and paragraph formatting in the recognized document. This option does not retain the original page layout.

For example, a multiple-column newspaper article would be formatted as one column in order of recognition instead as a page with side-by-side columns.



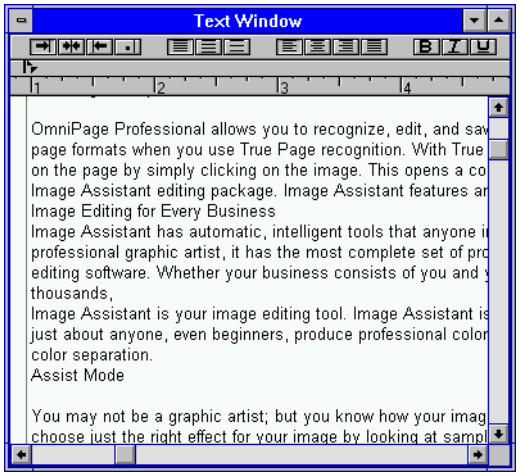
If you also selected *Retain Graphics*, any graphics in your document would appear at the bottom of the page. (You must also select *Multiple Columns* as the zoning method, or manually zone graphics if *One Zone* is the zoning method.)

This option attempts to reproduce the following during page recognition:

- Text alignment (justified, centered, or left-reading text)
- Font styles (select specific fonts in the Fonts settings panel)
- Font sizes
- Character styles (boldface, italics, underline)
- Paragraph groupings

Ignore All Formatting

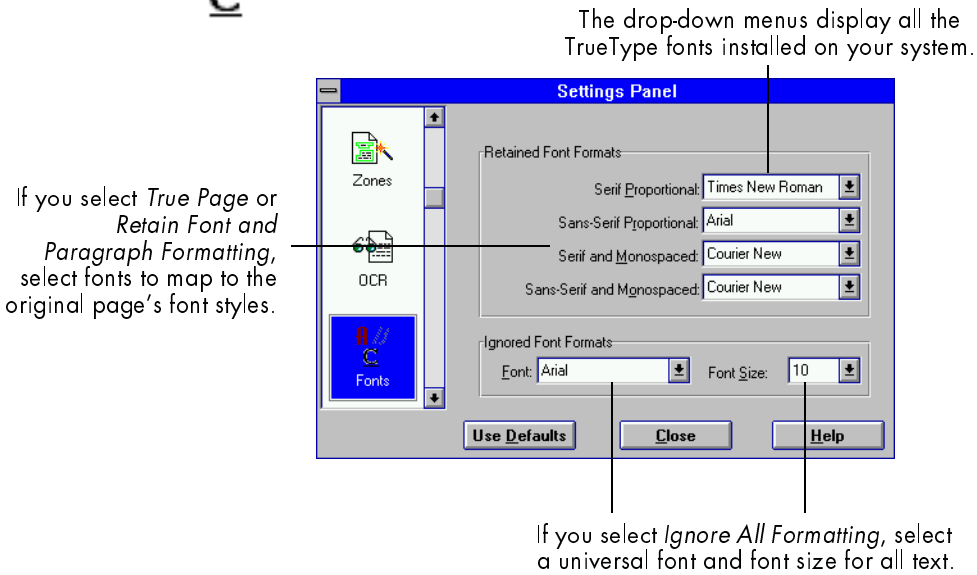
Select this to use one universal font and font size in the recognized document. Choose a font and font size for recognized text in the *Ignored Font Formats* section of the Fonts settings panel.



This option does not retain the original page layout. If you selected *Retain Graphics,* any graphics in your document would appear at the bottom of the page. This option is useful if you need only the text and or / graphics from a page. Recognition is speeded slightly if OmniPage does not have to map the fonts styles to those selected in the Fonts settings panel. See the next section for information on the Fonts settings panel.

Fonts Options

Click the Fonts icon in the Settings Panel to select options for retaining or ignoring the original font styles.



Retained Fonts Formats

You can select fonts to map to the various font styles in your document if you chose *True Page – Retain All Page Formatting* or *Retain Font and Paragraph Formatting* in the OCR settings panel.

There are four types of fonts available for mapping in this section. Each has the same drop-down list containing the TrueType fonts on your system. You can choose to map your fonts to the following font styles:

- *Serif Proportional*

  Character spacing varies depending on each character; short lines finish off the letter strokes. The body text in this manual is Palatino, a serif proportional font.

- *Sans-Serif Proportional*

  Character spacing varies depending on each character; letter strokes do not have finishing lines. The topic heads and subheads in this manual are Futura, a sans-serif proportional font.

- *Serif and Monospaced*

  Character spacing is the same for each character; short lines finish off the letter strokes. `Courier` is an example of this font type.

- *Sans-Serif and Monospaced*

  Character spacing is the same for each character; letter strokes do not have finishing lines. OCR-A is an example of this font type.

OmniPage detects font styles during recognition and formats recognized characters according to the font selected for that style. For example, if you selected *Times New Roman* in the *Serif Proportional* drop-down list, characters printed in Palatino (a serif proportional font used in this manual) would be reformatted as Times New Roman during recognition.

Ignored Font Formats

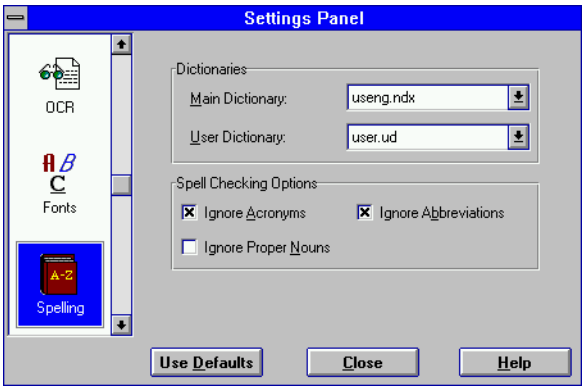
You must select a universal font and font size for recognized text in your document if you chose *Ignore All Formatting* in the OCR settings panel. Select a font in the *Font* drop-down list box and a font size in the *Font Size* drop-down list.

OmniPage ignores all font styles during recognition and formats recognized characters according to the font and font size you select.

For example, suppose you selected *Arial* in the *Font* drop-down list and *10* in the *Font Size* drop-down list. All characters, regardless of their current font and point size, would be changed to Arial 10-point text during recognition. The text would have no formatting (italic, bold, etc.) but paragraph order would be maintained.

Spelling Options

Click the Spelling icon in the Settings Panel to select dictionaries and spell-checking options.



Dictionaries

You can select one main dictionary and one user (personal) dictionary. The Language Analyst uses the selected dictionaries during recognition to check questionable words against dictionary entries. The dictionaries are also used when checking recognition. Always select the appropriate dictionaries for your document.

Main Dictionary

Select a main dictionary in the *Main Dictionary* drop-down list box. Main dictionaries have the file extension *.ndx.

OmniPage is delivered with the main dictionary or dictionaries appropriate to your country. To order dictionaries for additional languages, call your local Caere distributor or call Caere at (800) 535-SCAN.

User Dictionary

Select a user (personal) dictionary in the *User Dictionary* drop-down list box. User dictionaries have the file extension *.ud. Your default user dictionary is user.ud.

To create a new user dictionary or edit an existing user dictionary, choose *Edit User Dictionary...* in the Settings menu. For more information on creating and editing a user dictionary, please see “Edit User Dictionary” on page 149.

Spell-Checking Options

You can select the following spell-checking options:

Ignore Acronyms

OmniPage will ignore entirely capitalized words of four characters or less (for example, HUD, USDA).

Deselect *Ignore Acronyms* if you want the acronyms in your user dictionary to be checked or if you want to add acronyms to your user dictionary.

Ignore Proper Nouns

OmniPage will ignore a word not beginning a sentence that has a capitalized first letter (for example, in *He saw Althea throw...* OmniPage ignores the name Althea).

Deselect *Ignore Proper Nouns* if you want the proper nouns in your user dictionary to be checked or if you want to add proper nouns to your user dictionary.

Ignore Abbreviations

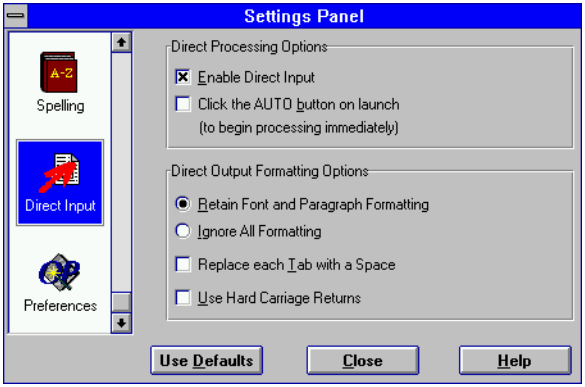
OmniPage will ignore a capitalized letter followed by three or fewer lowercase letters and a period (for example, Mrs., Dr., etc.).

Deselect *Ignore Abbreviations* if you want the abbreviations in your user dictionary to be checked or if you want to add abbreviations to your user dictionary.

You must select these options *before* performing recognition on a page or they will not take effect during the check recognition process. This is because OmniPage compiles its list of suspect word during OCR from the page being recognized.

Direct Input Options

Click the Direct Input icon in the Settings Panel to select processing and output options for Direct Input mode.



Use the Direct Input feature to initiate OCR processing from another application. See Chapter 5, Direct Input, for detailed information.

Direct Processing Options

You can select the following options under *Direct Processing Options:*

- *Enable Direct Mode*
- *Click the AUTO button on launch*

Enable Direct Mode

Select this option to enable Direct Input in registered applications. The *Direct Input...* command is placed in the File menu of an open, registered application when Direct Input mode is enabled. Deselect this option to remove the *Direct Input...* command from the File menu of a registered applications.

Click the AUTO button on launch

Select this option to begin automatic processing in Direct Input mode immediately when you choose *Direct Input...* in a registered application.

OmniPage uses the last-selected options when it opens Direct Input mode. Make sure the Settings Panel options and the choices in the process button drop-down lists are set correctly before you begin automatic processing.

Direct Output Options

You can select the following options under *Direct Output Formatting Options*:

- *Retain Font and Paragraph Formatting*
- *Ignore All Formatting*
- *Replace each Tab with a Space*
- *Use Hard Carriage Returns*

Retain Font and Paragraph Formatting

Select this to retain font and paragraph formatting in the recognized document. You can select fonts to map to the various font styles in your document in the *Retained Font Formats* section of the Fonts settings panel.

This option attempts to reproduce the following during page recognition:

- Text alignment (justified, centered, or left-reading text)
- Font styles (select specific fonts in the Fonts settings panel)
- Font sizes
- Character styles (boldface, italics, underline)
- Paragraph groupings

Ignore All Formatting

Select this to use one universal font and font size in the recognized document. Choose a font and font size for recognized text in the *Ignored Font Formats* section of the Fonts settings panel.

This option does not retain the original page layout. If you selected *Retain Graphics*, any graphics in your document appear at the bottom of the page. Recognition is speeded slightly if OmniPage does not have to map the fonts styles to those selected in the Fonts settings panel.

Replace each Tab with a Space

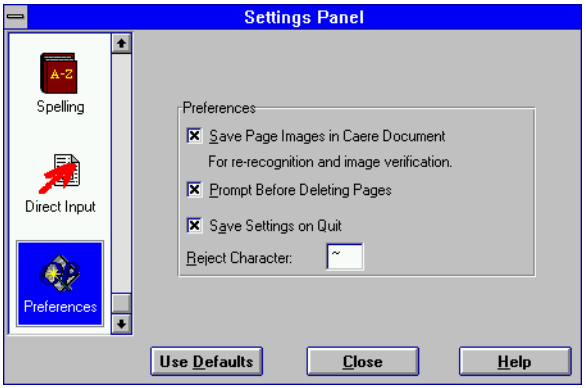
OmniPage replaces areas between text of approximately five or more spaces with a tab during recognition. Select this option to replace these areas with one space instead of a tab. Do not select this option if you want to maintain the tabular formatting in a table or similar document.

Use Hard Carriage Returns

Select this to place a return after each line of recognized text instead of wrapping each line.

Preferences Options

Click the Preferences icon in the Settings Panel to customize general OmniPage operations.



Save Page Images in Caere Document Format

Select this option to save original page images in Caere Documents. This image is the “picture” that appears in the zone window when you scan a page or load an image file.

To save the page image, make sure *Save Page Images in Caere Document* is selected *before* you scan or load a page image.

There are advantages to saving an image along with its recognized text as a Caere Document. You can always reopen a Caere Document and:

- use the Verification window to compare recognized text with the original page image
- defer recognition
- rezone and re-recognize pages at any time

This saves you the time of rescanning a page or reloading an image.

Saving a Caere Document without page images allows quicker processing and saves disk space but does not allow any of the above operations.

Prompt Before Deleting Pages

Select this if you want OmniPage to prompt you before carrying out the *Delete Page* command. This gives you the option of canceling the operation before deleting a page.

Save Settings on Quit

Select this if you want to save the current OmniPage settings automatically when you exit the program. The Settings Panel options and language selection(s) will be retained until you select new settings. You must close the Settings Panel before you exit OmniPage, however, or it will revert to the last-selected options.

Reject Character

Unrecognizable characters are represented by a red reject character in the recognized document. In the *Reject Character* edit box, type in any character that you want to be the reject character. The default character is a tilde (~).

For example, if OmniPage could not recognize the J in REJECT, and the tilde (~) was the reject character, the string RE~ECT would appear in your recognized document.

Chapter 5

Direct Input

This chapter explains how to initiate OCR processing from an open application and paste recognized text directly from OmniPage into that application. OmniPage has a special Direct Input mode that can be initiated from any compatible application.

Most commands and settings in Direct Input mode are the same as those found in the regular OmniPage mode. This chapter describes the differences. Refer to Chapter 3, Commands and Settings, and Chapter 4, The Settings Panel, for detailed information on commands and settings not discussed in this chapter.

This chapter contains the following topics:

- Registering Applications
- Using Direct Input from Another Application
- Selecting Settings for Direct Input

Registering Applications

An application must be registered before it can be used to access Direct Input mode. A variety of applications are compatible with Direct Input.

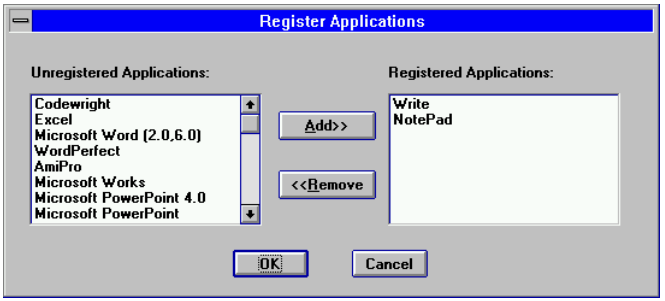
Registering an Application

1 Launch OmniPage if it is not already open.

2 Choose *Register Applications...* in the Settings menu.

This command is enabled only when *Enable Direct Input* is selected in the Direct Input settings panel.

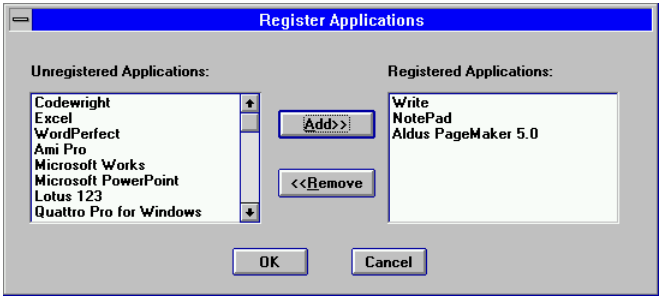
The Register Applications dialog box appears. It displays a list of all applications that are compatible with Direct Input. The Windows programs Write and NotePad are pre-registered.



3 Select an application to register in the *Unregistered Applications* list box.

4 Click *Add>>*.

The application moves into the *Registered Applications* list box.



5 Select and move as many applications as you like.

6 Click *OK* when you are done.
OmniPage places the *Direct Input...* command in the File menu of the registered application(s) above the *Exit* command.

The dialog box displays a list of *compatible* applications only. The list of available applications is static; you cannot add applications to the list. An application must be installed in order to use it with Direct Input.

Remove an Application From the Registered List

Removing an application that you do not need from the *Registered Applications* list box frees up a small amount of memory.

1 Choose *Register Applications...* in the Settings menu to open the Register Applications dialog box.

2 Select an application in the *Registered Applications* list box.

3 Click *<<Remove.*
The application name moves into the *Unregistered Applications* list box.

4 Select and remove as many applications as you like.

5 Click *OK* when you are done.
OmniPage removes the *Direct Input...* command from the File menu of the application(s) you selected.

Using Direct Input from Another Application

OmniPage places a *Direct Input...* command in an application's File menu above the *Exit* command once that application is registered. Choose *Direct Input...* to initiate Direct Input processing.

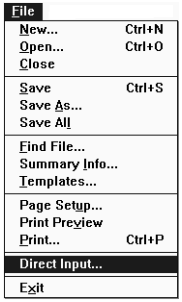
Some applications, such as Word and Notepad, allow you to launch multiple copies of the application. The *Direct Input...* command only appears in the first copy of the application launched.

Accessing OmniPage in Direct Input Mode

1 Align your document(s) in your scanner or ADF if you plan to scan.

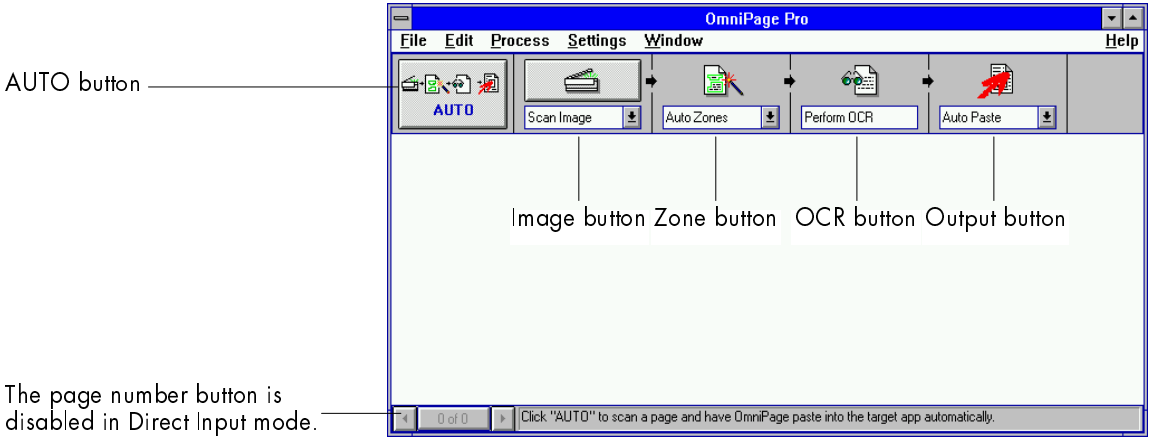
2 Open or switch to any registered application.
Microsoft Word is used in this example.

3 Choose *Direct Input...* in the program's File menu.



- OmniPage launches in Direct Input mode if it was closed.
- OmniPage switches to Direct Input mode if it was already open or was running as a desktop icon.
  If a document is open in OmniPage at the time, a dialog box prompts you to close and save the OmniPage document before starting Direct Input.

The Direct Input window appears.



Automatic processing begins immediately if you had selected *Click the AUTO button on launch* in the Direct Input settings panel before initiating Direct Input.

4 Select the appropriate process button settings and Settings Panel options for your document if you did not selected automatic processing to begin immediately.

5 Click *AUTO* or the Image button to begin processing.

OmniPage processes the page according to your settings. After processing, the Direct Input window closes and OmniPage returns to the state it was in before Direct Input mode began: open, iconized, or closed.

See the next section for more information on how your document is processed in Direct Input mode.

Direct Input Mode Processing

What OmniPage does after the Direct Input window opens depends on the settings you selected.

See Chapter 3, Commands and Settings, and Chapter 4, The Settings Panel, for detailed information on how your settings affect OCR output.

Acquiring an Image When no Image is Open

Automatic processing begins immediately if you selected *Click the AUTO button on launch* in the Direct Input settings panel. Otherwise, you must click either *AUTO* to begin automatic processing or the Image button to acquire an image.

- OmniPage begins scanning if *Scan Image* or *Scan Multiple* is selected under the Image button.
- OmniPage opens the Load Image dialog box if *Load Image* is selected under the Image button.
- You cannot open Caere Documents (`*.met`) in Direct Input.

Acquiring an Image When Another Image is Open

Unlike regular OmniPage mode, Direct Input only appends a newly loaded or scanned document to the current open document if that document has been recognized.

- If you scan or load an image but do not recognize it, and then click the Image button, OmniPage warns that the image will be replaced. The same is true for an image that has been zoned but not recognized.

- If you scan or load an image, recognize it, and then click the Image button, OmniPage automatically appends the newly loaded or scanned image to the currently open document.

  The page number button is disabled in Direct Input mode so OmniPage appends the new image to the last processed page.
- You cannot open Caere Documents (`*.met`) in Direct Input.

Zoning

There are at least two options under the Zone process button, *Auto Zones* and *Manual Zones*. There may also be zone templates if you have created and saved any. See "Save Zone Template" on page 100.

The Zone button is active when either *Auto Zones* or a specific zone template is selected.

- OmniPage automatically draws zones if *Auto Zones* or a zone template is selected and you clicked *AUTO*.
- You must click the Zone button after scanning or loading an image if you selected *Auto Zones* or a zone template but did not click *AUTO*.
- You can draw zones manually after an image is loaded or scanned if *Manual Zones* is selected under the Zone button.

  Remember that Direct Input does not save graphics. Graphic zones are ignored during processing.

OCR

You must click the OCR button after zoning is done unless automatic processing is in progress. OmniPage then recognizes the text according to the way it was zoned and what OCR settings panel options were selected. Unlike in regular OmniPage mode, recognized text is not displayed in a text window. You can see the text after it is pasted into your open application or copied to the Clipboard.

Output

You must click the Output button after OCR is done unless automatic processing is in progress.

- OmniPage pastes the recognized text directly into your open, registered application if *Auto Paste* is selected under the Output button.
- OmniPage copies the recognized text to the Clipboard if *Copy to Clipboard* is selected under the Output button. You can then paste the text into any application that supports the Clipboard.

Active Process Buttons

You can click any button that is active (raised instead of flat) to start processing at any point not yet done

For example, if you chose *Load Image*, *Auto Zones*, *Perform OCR*, and *Auto Paste* under the process buttons and then clicked the Image button. All four buttons are active after the image is loaded.

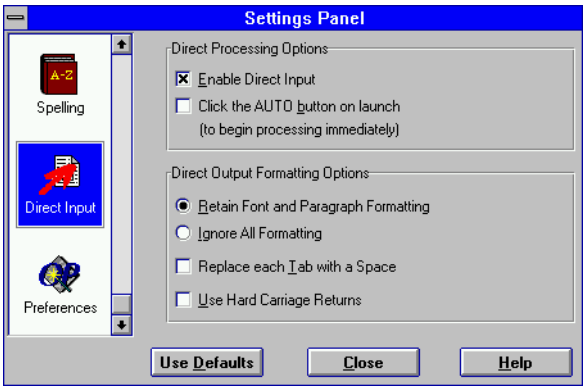
You could click the Output button and OmniPage would zone and recognize the image and then paste the recognized text into your open application.

Selecting Settings for Direct Input

It is always important to select the right settings before processing. Use the Settings Panel, the toolbar, and the menu items to set your processing options before scanning a page or loading an image.

The Direct Input Settings Panel

Choose *Settings Panel...* in the Settings menu or the keyboard shortcut Ctrl-E to open the Settings Panel. (There are no shortcut command buttons in the Direct Input mode as there are in OmniPage mode.)



Please note the following:

- If you deselect *Enable Direct Input*, the *Register Applications...* command in the Settings menu is disabled in the regular OmniPage mode.
- If you deselect *Enable Direct Input*, the *Direct Input...* command in any registered application's File menu disappears immediately.

You can continue processing if you are already in Direct Input mode at the time.

- If you select *Click the AUTO button on launch*, Direct Input processing begins as soon as the Direct Input window opens, the same as if you had clicked the *AUTO* button.

  The last-selected Settings Panel options and the last-select Direct Input process commands are used.

See "Direct Input Options" on page 177 for detailed information on Settings Panel options for Direct Input mode.

Note that the following Settings Panel options are disabled in Direct Input mode.

Scanner Options

- *Scan Until Empty*

  Select *Scan Multiple* in the Image button drop-down list in Direct Input mode instead if you want to scan multiple pages.

- *Double-sided pages*

  Direct Input mode does not support recognition of double-sided pages. Launch the regular OmniPage program to process double-sided pages.

OCR Options

- *Retain Graphics*

  Direct Input mode ignores graphics. Use the regular OmniPage mode if you want to save graphics on a page.

- *Output Formatting Options*

  Use the Direct Input settings panel to set output formatting options such as whether to retain font and paragraph styles.

Spelling Options

- *Spell Checking*

  These options are not available because Direct Input does not display recognized text window. It pastes it directly to your target application or copies it to the Clipboard. You can check recognition in your target application.

Preferences Options

- *Save Page Images in Caere Document*

  Direct Input mode does not save page images. Use the regular OmniPage mode for this.

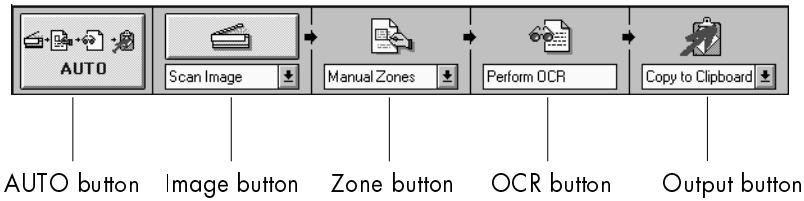
- *Prompt Before Deleting Pages*

  You cannot delete pages in Direct Input mode.

See Chapter 4, The Settings Panel, for information on other Settings Panel options.

The Direct Input Toolbar

The Direct Input toolbar has an extra process button and no shortcut command buttons.



Most of its functions are the same as those in the regular OmniPage mode. Differences are explained in this section. See “The Toolbar” on page 82 for detailed information.

AUTO Button

This button has the same function as it does in the OmniPage mode. It initiates automatic processing when clicked. It uses the selected options in each process button to process a document.

Image Button

Select *Scan Multiple* in Image button drop-down list to scan more than one page with automatic processing. (This command is not available in the regular OmniPage mode.) All pages are scanned if you have a scanner with an automatic document feeder (ADF) and use automatic processing. You will receive a prompt to place another page in the scanner if you use automatic processing but do not have an ADF.

Zone Button

This button has the same function as it does in the OmniPage mode. OmniPage automatically draws zones if you select *Auto Zones* and use automatic processing. You can also click the Zone button after an image is

loaded or scanned. OmniPage stops processing for you to draw zones if you select *Manual Zones*.

OCR Button

There is no drop-down list under this button. The only setting is *Perform OCR*. Click this button to perform OCR on a zoned page. OCR begins automatically if you click the *AUTO* button and *Auto Zones* is also selected in the Zone button's drop-down list.

Process your document or image file in the regular OmniPage mode if you need to use either the *Train OCR* or *Defer OCR* command.

Output button

Use this button to choose a destination for your recognized text.

- Select *Auto Paste* to paste recognized text directly into your open registered application after OCR.
- Select *Copy to Clipboard* to copy recognized text to the Clipboard after OCR. OmniPage does this automatically if you close a registered application or if there is no document open in the application when the Output button is clicked.

Select the correct process button settings and Settings Panel options for your document before beginning automatic processing. Automatic processing begins immediately in Direct Input mode if *Click the AUTO button on launch* is selected in the Direct Input settings panel. Select settings in Direct Input before you exit if you want to use those same settings with the automatic processing on launch option.

The Direct Input Menus

The Direct Input mode window contains many of the same menus and commands as the regular OmniPage mode:

- File menu
- Edit menu
- Process menu
- Settings menu
- Windows menu

This section describes commands found only in Direct Input mode. See Chapter 3, Commands and Settings, for a description of all other menu commands.

Process Menu

See "The Process Menu" on page 117 for a description of other commands.

Auto Paste

Choose *Auto Paste* to paste recognized text directly into your open registered application after OCR. This command performs the same function as the Output button when *Auto Paste* is selected in its drop-down list.

Copy to Clipboard

Choose *Copy to Clipboard* to copy recognized text to the Clipboard after OCR. OmniPage does this automatically if you close a registered application or if there is no document open in the application when the Output button is clicked.

This command performs the same function as the Output button when *Copy to Clipboard* is selected in its drop-down list.

Window Menu

Choose *Exit Direct* to return OmniPage to whatever state it was in when Direct Input mode began: iconized, running, or closed. Any open image or recognized text is discarded when you choose this menu command. You cannot save documents or export images in Direct Input. See "The Window Menu" on page 152 for a description of other commands.

Chapter 6

Using True Page

Ordinarily, you use OCR to turn a printed page into editable electronic text. Regardless of the original page layout, the resulting text appears in a single column. You then save this text in a file format suitable for your target application (such as a word processor) where you apply your own page layout and formatting.

True Page offers a solution for retaining the original look of a page during the OCR process. In addition to turning images into text, True Page tries to preserve font and paragraph formatting and the positioning of text and graphics on a page.

True Page handles documents differently depending on the type of page you process and the Settings Panel options you select. To get the most benefit out of True Page, learn how to identify your page type and select the appropriate settings for it.

Please read this chapter for information about:

- How True Page Works
- Page Types
- Settings Panel Options for True Page
- Target Applications for True Page

How True Page Works

True Page tries to replicate the look of a page by analyzing and retaining its formatting elements during recognition. True Page retains page margins, column structure, paragraph structure, tab formatting, indentation, line spacing, font attributes, and positioning of graphics.

True Page handles documents according to the options you select in the Settings Panel. The zoning method you select is particularly important:

- *Multiple Columns*

If you select this as your zoning method, True Page uses *frames* to preserve the positioning of text and graphic elements on the page. Frames are formatting boxes that are used for designing page layout; they can contain text or graphics. For example, every paragraph in a document may be contained within a separate frame. You can resize frames and move them around to modify your document's page layout.

- *Single Column or Table* or *One Zone*

  If you select one of these as your zoning method, True Page does not use frames to preserve the positioning of text and graphic elements on the page.

After OCR with True Page, text and graphics appear in frames in the OmniPage text window regardless of the zoning method used. However, if the *Single Column or Table* or *One Zone* zoning method was used, frames will be removed when you save a document to another file format.

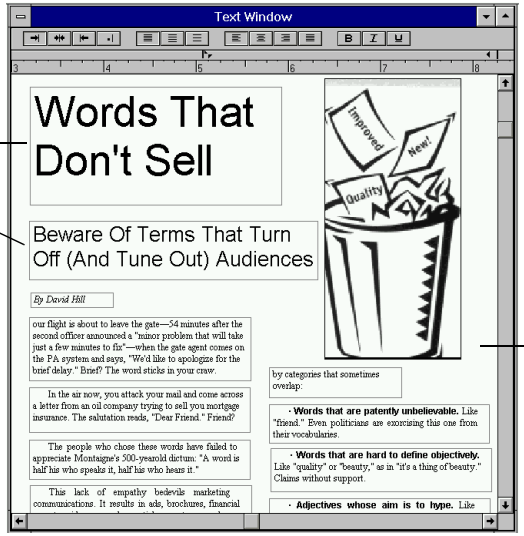
The next sections explain the difference between True Page output with frames and without frames.

True Page With Frames

The best way for True Page to preserve side-by-side columns is to create frames around areas of text and graphics. Therefore, frames are automatically used when you select the *Multiple Columns* zoning method with True Page. These frames are exported intact when you save your document to a file format that supports frame formatting.

After recoanition with True Page, a document is formatted in frames in the text window.

If the *Multiple Columns* zoning method was used, frames are exported intact when you save to an appropriate file format.



You can resize and move frames around in the OmniPage text window. See "Working With Frames" on page 35 for more information.

When to Use True Page for Multiple-Column Pages

Use True Page for multiple-column pages when you want to retain the appearance (formatting and layout) of a page and do not plan to edit or reformat the page extensively. Do not use True Page if you plan to apply a lot of your own page formatting and layout. It is easier to edit a page and apply new formatting when it has fewer formatting attributes.

For example, if you wanted to incorporate a newspaper article into your own newsletter, you would perform OCR on the text and graphic elements of the article and then put this information into your own page layout. To achieve this, you would select one of the other OCR output options, *Retain Font and Paragraph Formatting* or *Ignore All Formatting* with the *Multiple Column* zoning method. You would not use True Page because it would apply extra, unnecessary formatting. However, if you want to replicate a newspaper article as closely as possible without having to design your own layout, you would use True Page.

For examples of multiple-column pages, see "Multiple-Column Documents" on page 197.

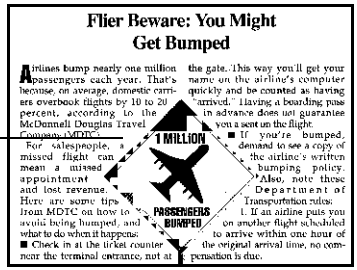
Factors That Influence Frame Formatting

Consider these factors before you choose True Page formatting for multiple-column pages:

- Pages with rectangularly shaped text and graphic areas and non-stylized fonts are the best candidates for True Page formatting.

  Highly stylized formatting is difficult to replicate especially when you save recognized documents to other file formats.

Non-rectangular text formatting around a graphic is difficult to replicate



Flier Beware: You Might Get Bumped

Airlines bump nearly one million passengers each year. That's because, on average, domestic carriers overbook flights by 10 to 20 percent, according to the McDonnell Douglas Travel Company (MDTC).

For salespeople, a missed flight can mean a missed appointment and lost revenue. Here are some tips from MDTC on how to avoid being bumped, and what to do when it happens:

■ Check in at the ticket counter near the terminal entrance, not at the gate. This way you'll get your name on the airline's computer quickly and be counted as having "arrived." Having a boarding pass in advance does not guarantee you a seat on the flight.

■ If you're bumped, demand to see a copy of the airline's written bumping policy. Also, note these Department of Transportation rules: 1. If an airline puts you on another flight scheduled to arrive within one hour of the original arrival time, no compensation is due.

1 MILLION PASSENGERS BUMPED

- There should be adequate white space between columns, paragraphs, and graphics.

  Glance quickly at your page; if separate areas are not immediately obvious, then you probably should draw zones manually to distinguish these separate areas.

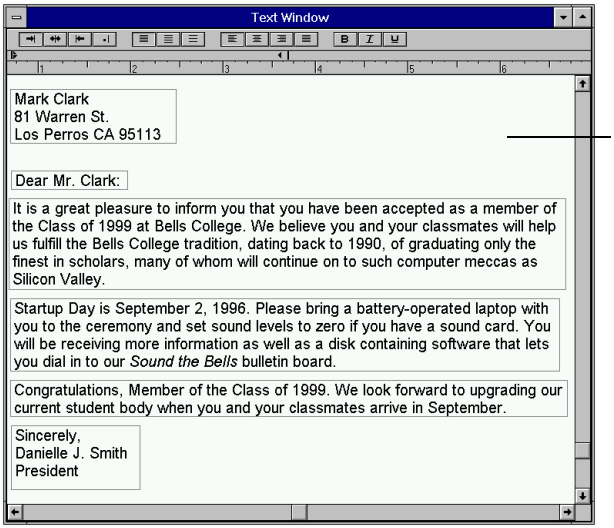
- True Page is effective for multiple-column documents only when you save your document to a True Page-compatible application.

  Your application must support frame-based layouts. Even though a page may appear correctly in OmniPage's text window, you will run into difficulty if you save it to an application that is not designed to handle frames. See "Target Applications for True Page" on page 205 for more information.

True Page Without Frames

Frames are unnecessary to preserve single-column formats. Tabs (or spaces) are used to preserve side-by-side formatting in tables. Therefore, frames are not used when you select the *Single Column or Table* zoning method with True Page.

After recognition with True Page, a document is formatted in frames in the text window.



You can resize and move frames around in the OmniPage text window. See "Working With Frames" on page 35 for more information.

If the *Single Column or Table* or *One Zone* zoning method was used, frames are not retained when you save to another file format.

When to Use True Page for Single-Column Pages or Tables

Use True Page for single-column documents or tables if your OCR goal is to get editable text and retain as much formatting as possible so you do not have to apply your own. True Page retains more formatting elements than the *Retain Font and Paragraph Formatting* and *Ignore All Formatting* options.

However, if your OCR goal is to get editable text and then extensively format your document's format and layout in another application, select the *Ignore All Formatting* or *Retain Font and Paragraph Formatting* option. It is easier to edit a page and apply new formatting when it has fewer formatting attributes.

For examples of single-column pages and tables, see "Single-Column Documents" on page 198 and "Tables" on page 199.

Page Types

The quality and layout of your original page influence the way True Page reproduces its appearance.

These factors influence True Page output:

- Accurate text recognition is essential to duplicating the original page's appearance successfully.

  For accurate text recognition, the print on the page should be reasonably clean, crisp, and free of notes, lines, or doodles. Characters must be distinct: separated from each other and not blotched together or overlapping.

- Highly stylized formatting is difficult to replicate.

  Pages with rectangular text and graphic blocks and non-stylized fonts are the best candidates for True Page formatting.

- TrueType fonts substituted by True Page may not take up the same amount of space that your original fonts did, so line and paragraph lengths may vary.

- True Page results may vary for different page types.

  Experiment with various settings to get the best results for your particular page.

The next sections describe different types of pages and recommended settings for True Page output.

Multiple-Column Documents

Newspaper and magazine articles are examples of multiple-column documents. If you use True Page for these types of documents, select the *Multiple Columns* zoning method for best results.

True Page tries to replicate the font attributes, paragraph structure, column structure, and text and graphic positioning of the original page. The document is formatted in frames in the OmniPage text window. These frames are exported intact when you save the document to an appropriate file format.

The *Multiple Columns* zoning method tries to identify and retain graphics on a page when you select *Retain Graphics* in the Settings Panel. If graphics are missed during OCR, try drawing zones manually and identifying graphics with the *Graphic* zone contents file.

Multiple-column pages can be challenging to replicate because they often use stylized fonts, a tight column structure, and non-rectangular paragraph and graphic layouts. To get the best results, you may want to experiment with different settings.

Example — Multiple-Column Page

Select *Multiple Columns* as the zoning method.

Select TrueType fonts in the Fonts settings panel that are the closest match to the fonts on your page.

It May Not Be Pretty...

...but this book is so practical it will likely become a dog-eared favorite in your resource library. "How To Write Better Business/Industrial Copy" by Bernie Powers, wastes no time in providing tools that make significant improvements in business-to-business and industrial copy. And it's packed with lots of concrete examples and checklists to help you apply these concepts to your own marketing communications, including:

- The seven key differences between business/industrial and consumer copywriting — and their importance in designing effective marketing communications materials.
- A six-step formula for writing better industrial copy.
- Nine ways to make copy more readable.
- How to write effective headlines, subheads, captions and closings.
- Specific tips for creating print ads, brochures, PR materials and audio-visual scripts.
- 37 ways to use value added marketing to write copy that forcefully differentiates your product from your competition's.
- 333 *examples of effective headlines.*
- Two dozen questions to ask before you begin writing.

Bernie Powers is a 20-year veteran of business-to-business advertising whose work as appeared in trade journals worldwide. He has won numerous readership studies for a variety of business and industrial clients, and is considered to be one of the world's foremost experts in writing business-to-business and industrial copy. ❖

For more information or to order BMA's new book, contact the Business Marketing Association at 800/664-4262 (150-pages, softcover; $19.95).

Words That Don't Sell

Beware Of Terms That Turn Off (And Tune Out) Audiences

By David Hill

Your flight is about to leave the gate — 54 minutes after the second officer announced a "minor problem that will take just a few minutes to fix" — when the gate agent comes on the PA system and says, "We'd like to apologize for the brief delay." Brief? The word sticks in your craw.

In the air now, you attack your mail and come across a letter from an oil company trying to sell you mortgage insurance. The salutation reads, "Dear Friend." Friend?

The people who chose these words have failed to appreciate Montaigne's 500-year-old dictum: "A word is half his who speaks it, half his who hears it."

This lack of empathy bedevils marketing communications. It results in ads, brochures, financial reports, videos, speeches, articles, posters, employee bulletins, etc., that fail to connect with their intended audiences. Aiming to engage, these communications do the opposite, often simply because of an insensitive choice of words.

When you're writing sales pitches (i.e., any communication designed to persuade recipients to think/feel/act a certain way), it's important to recognize two things: 1) words that work when you're face-to-face and one-on-one may fall flat in a speech, ad or brochure; and 2) there is no such thing as a "throwaway word." Even apparently innocuous words can close your listeners' and readers' minds instantly and irrevocably.

Terms That Turn Off

What are these "turn off" words? Here's a list of some of the worst offenders, grouped by categories that sometimes overlap:

- **Words that are patently unbelievable.** Like "friend." Even politicians are eschewing this one from their vocabularies.
- **Words that are hard to define objectively.** Like "quality" or "beauty," as in "it's a thing of beauty." Claims without support.
- **Adjectives whose aim is to hype.** Like "exciting," "wonderful," "spectacular," "terrific" or "great."
- **Unsupported superlatives and comparatives.** Like "lowest rates in town" or "second to none," both of which are read or heard as, "same as everybody else."
- **Adverbs whose aim is to temper.** Like the aforementioned "briefly" or those old favorites, "very," "really," "virtually" and "significantly."
- **Buzz words.** Like "world class" and "our people."
- **Any combination of the above.** Like "world class quality that's really spectacular, thanks to our dedicated people."

Words like these are like a tap on your knee. They trigger a reflexive reaction, usually unconscious. They prompt audiences to tune out, turn off and disconnect.

Among words that disconnect, one merits special attention, a paragraph of its own, because it's especially powerful and ubiquitous. We're inclined to use it whenever we have something we really want to tell our audiences. The word is "We."

"We" forewarns audiences that what's

(continued on page 4)

If you draw zones manually, identify graphics with the *Graphic* zone contents file.

For tight spacing between columns, try drawing zones manually around each column. Order zones carefully.

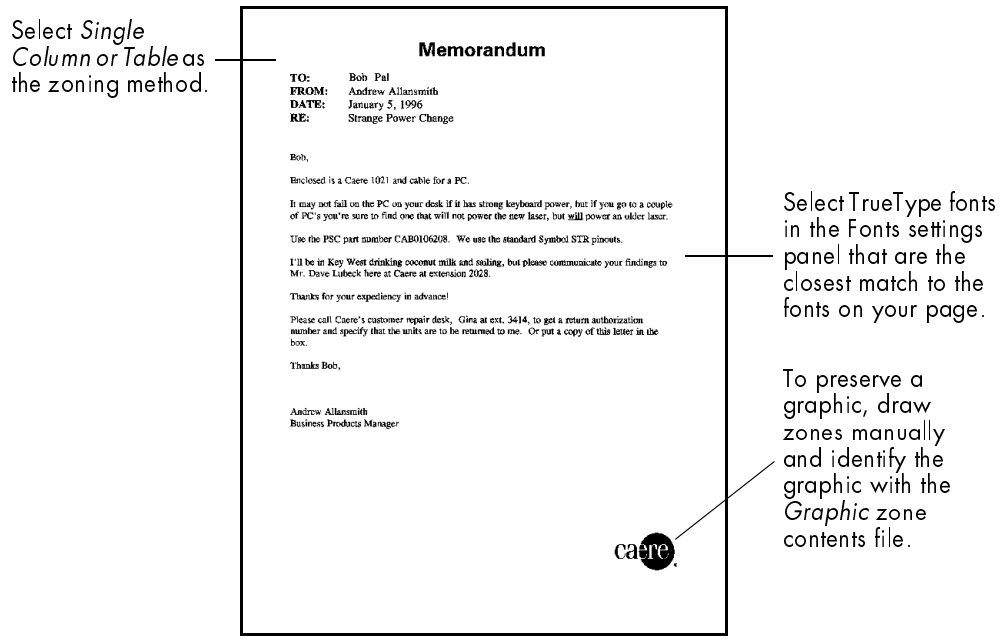
Single-Column Documents

Memos and letters are examples of single-column documents. If you use True Page for these types of documents, select the *Single Column or Table* zoning method for best results.

True Page tries to replicate the font attributes, paragraph structure, and text and graphic positioning of the original page. Although the document is formatted in frames in the OmniPage text window, these frames are not retained when you save the document to another file format.

The *Single Column or Table* zoning method retains graphics on a page only if you select *Retain Graphics* in the OCR settings panel, draw zones manually, and identify any graphics with the *Graphic* zone contents file.

Example — Single-Column Page



Tables

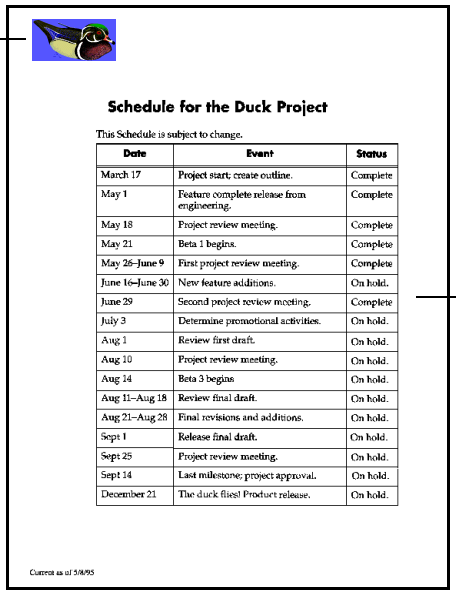
Tables, such as spreadsheets, have information set up in rows and columns. The columnar information in tables needs to remain side by side. If you use True Page for these types of documents, select the *Single Column or Table* zoning method for best results.

True Page tries to replicate the font attributes, table structure, and other text and graphic positioning of the original page. Tabs are used to preserve side-by-side formatting in tables (spaces are used if OmniPage detects less than five spaces between columns). Although the document is formatted in frames in the OmniPage text window, these frames are not retained when you save the document in another file format.

The *Single Column or Table* zoning method retains graphics on a page only if you select *Retain Graphics* in the OCR settings panel, draw zones manually, and identify any graphics with the *Graphic* zone contents file.

Example — Table

To preserve a graphic, draw zones manually and identify the graphic with the *Graphic* zone contents file.



Schedule for the Duck Project

This Schedule is subject to change.

Date	Event	Status
March 17	Project start; create outline.	Complete
May 1	Feature complete release from engineering.	Complete
May 18	Project review meeting.	Complete
May 21	Beta 1 begins.	Complete
May 26–June 9	First project review meeting.	Complete
June 16–June 30	New feature additions.	On hold.
June 29	Second project review meeting.	Complete
July 3	Determine promotional activities.	On hold.
Aug 1	Review first draft.	On hold.
Aug 10	Project review meeting.	On hold.
Aug 14	Beta 3 begins	On hold.
Aug 11–Aug 18	Review final draft.	On hold.
Aug 21–Aug 28	Final revisions and additions.	On hold.
Sept 1	Release final draft.	On hold.
Sept 25	Project review meeting.	On hold.
Sept 14	Last milestone; project approval.	On hold.
December 21	The duck flies! Product release.	On hold.

Current as of 5/6/95

Select *Single Column or Table* in the Zones settings panel.

Combination Documents

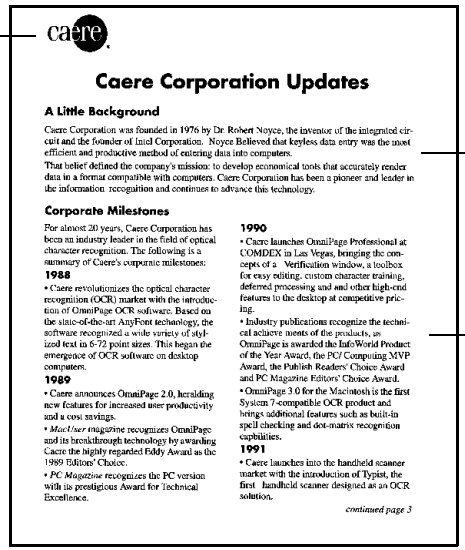
Combination documents have different types of column layouts on the same page. For example, a newsletter could have a mixture of one-column and multiple-column layouts. If you use True Page for combination documents, select the *Multiple Columns* zoning method for best results.

True Page tries to replicate the font attributes, paragraph structure, column structure, and text and graphic positioning of the original page. The document is formatted in frames in the OmniPage text window. These frames are exported intact when you save the document to an appropriate file format.

Since combination pages can be complex, experiment with different settings to get the best results. For example, try drawing zones manually if auto zoning does not work well to retain all elements on a page.

Example — Combination Page

If auto zoning misses a graphic, draw zones manually and identify the graphic with the *Graphic* zone contents file.



caere

Caere Corporation Updates

A Little Background

Caere Corporation was founded in 1976 by Dr. Robert Noyce, the inventor of the integrated circuit and the founder of Intel Corporation. Noyce Believed that keyless data entry was the most efficient and productive method of entering data into computers.

That belief defined the company's mission: to develop economical tools that accurately render data in a format compatible with computers. Caere Corporation has been a pioneer and leader in the information recognition and continues to advance this technology.

Corporate Milestones

For almost 20 years, Caere Corporation has been an industry leader in the field of optical character recognition. The following is a summary of Caere's corporate milestones:

1988

• Caere revolutionizes the optical character recognition (OCR) market with the introduction of OmniPage OCR software. Based on the state-of-the-art AnyFont technology, the software recognized a wide variety of stylized text in 6-72 point sizes. This began the emergence of OCR software on desktop computers.

1989

• Caere announces OmniPage 2.0, heralding new features for increased user productivity and a cost savings.

• *MacUser* magazine recognizes OmniPage and its breakthrough technology by awarding Caere the highly regarded Eddy Award as the 1989 Editors' Choice.

• *PC Magazine* recognizes the PC version with its prestigious Award for Technical Excellence.

1990

• Caere launches OmniPage Professional at COMDEX in Las Vegas, bringing the concepts of a Verification window, a toolbox for easy editing, custom character training, deferred processing and and other high-end features to the desktop at competitive pricing.

• Industry publications recognize the technical achievements of the products, as OmniPage is awarded the InfoWorld Product of the Year Award, the PC/ Computing MVP Award, the Publish Readers' Choice Award and PC Magazine Editors' Choice Award.

• OmniPage 3.0 for the Macintosh is the first System 7-compatible OCR product and brings additional features such as built-in spell checking and dot-matrix recognition capabilities.

1991

• Caere launches into the handheld scanner market with the introduction of Typist, the first handheld scanner designed as an OCR solution.

continued page 3

One-column format.

Two-column format; select *Multiple Columns* in the Zones settings panel.

Settings Panel Options for True Page

The Settings Panel options you select influence the way True Page reproduces a document's appearance. Since each page varies, experiment with different settings to get the best True Page results.

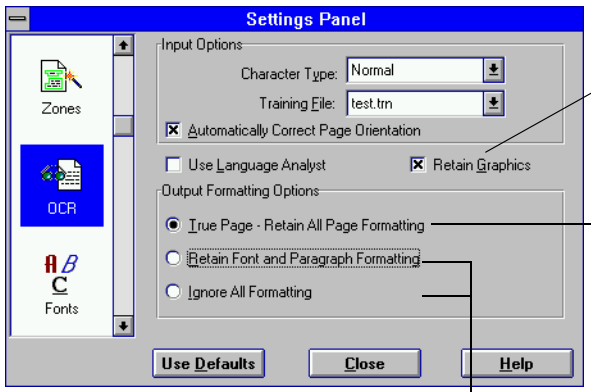
Accurate text recognition is the first step in successfully duplicating your page's appearance. For every page that you recognize, choose the settings that will achieve the most accurate results. See Chapter 4, The Settings Panel, for more information.

The Settings Panel options that are particularly important to True Page output include:

- OCR Options
- Zones Options
- Fonts Options

OCR Options

Select *True Page — Retain All Page Formatting* in the OCR settings panel to designate True Page output for your document. True Page tries to replicate a document's page margins, column structure, paragraph structure, tab formatting, indentation, justification, line spacing, font attributes, and positioning of graphics.



Select *Retain Graphics* to preserve graphics on the page.

Select *True Page* to retain font and paragraph formatting and the positioning of text and graphics.

The other *Output Formatting Options* retain less formatting than True Page and do not try to retain the original positioning of text and graphics.

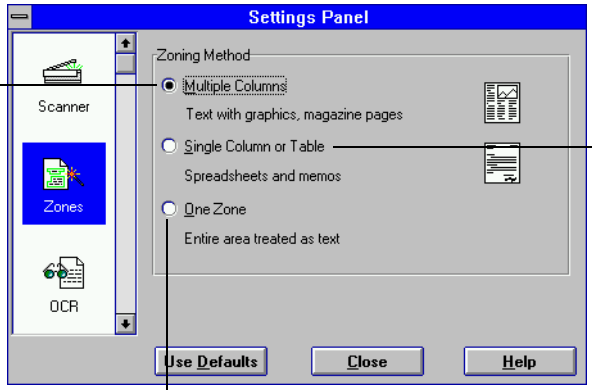
The two other formatting options are easier to edit because they add fewer formatting attributes than True Page. The *Retain Font and Paragraph Formatting* option tries to replicate a document's font attributes and paragraph structure but does not reproduce the other formatting elements such as positioning of graphics and columns. The *Ignore All Formatting* option applies minimal formatting to a document during recognition.

Zones Options

Zones created on the original page image influence True Page output. For example, if you draw zones manually and you do not include a paragraph in any zone, that paragraph will be skipped entirely by the OCR process and the recognized page will not look like the original.

The flow of text on a page is determined by the zoning method you select in the Zones settings panel regardless of whether you draw zones manually or OmniPage creates zones automatically. For True Page output, select either *Multiple Columns* or *Single Column or Table* for best results.

Multiple Columns looks for text that is formatted into several columns. For True Page formatting, frames are used to retain a page's column and paragraph structure.



Single Column or Table looks for one-column text formatting or side-by-side table formatting. For True Page formatting, frames are not used.

One Zone looks for text across the entire page. For True Page formatting, frames are not used.

Be sure to select *Retain Graphics* in the OCR settings panel to preserve any graphics on your page. If you use the *Single Column or Table* or *One Zone* zoning method, you must draw zones manually and identify the graphic with the *Graphic* zone contents file. You should also try this if you have a complex page with graphics intermixed with text and auto zoning does not identify all of the graphics correctly.

Fonts Options

The font choices that you make in the Settings Panel play a big role in achieving the appearance of your original page. For True Page output, OmniPage determines the style of fonts on a page and maps them to the TrueType fonts you select in the Fonts settings panel.

Be sure that your system's TrueType fonts are enabled. The more TrueType fonts you have installed on your system, the more options you have for mapping.

There are four basic styles that OmniPage looks for on a page:

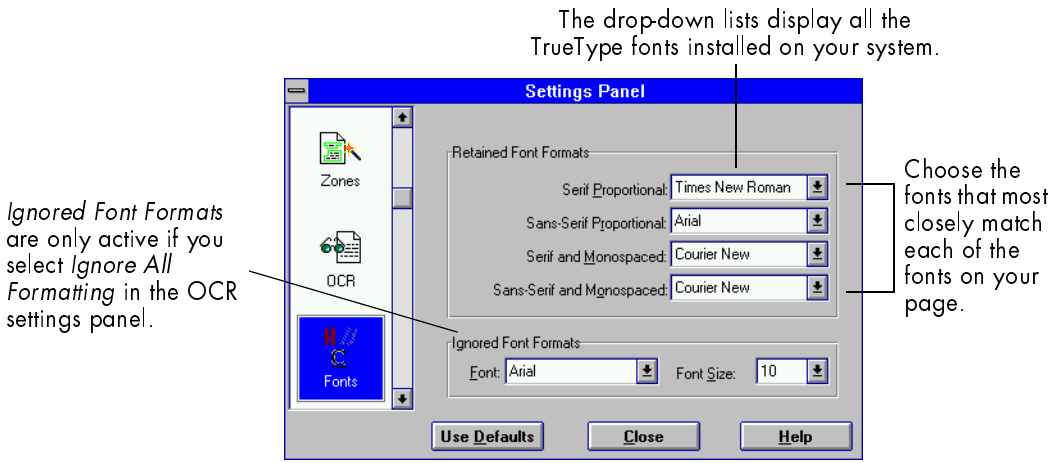
- Serif Proportional — short lines finish off the letter strokes; character spacing varies depending on each character.

  For example: Times New Roman
- Sans Serif Proportional — letter strokes do not have finishing lines; character spacing varies depending on each character.

  For example: Arial
- Serif and Monospaced — short lines finish off the letter strokes; character spacing is equal for each character.

  `For example: Courier`
- Sans Serif and Monospaced — letter strokes do not have finishing lines; character spacing is equal for each character.

  `For example: OCRA`



For example, if your document has text formatted with Times Roman and Helvetica fonts, select *Times Roman* as the Serif Proportional font and *Helvetica* as the Sans Serif Proportional font.

True Type fonts that are not an exact match for fonts in your document may not take up the same amount of space especially if the original fonts are stylized. For example, magazines often use titles with narrow, stylized fonts that are difficult to match with basic TrueType fonts. This may result in variations in line and paragraph lengths.

Target Applications for True Page

The formatting that is retained when you save your recognized document in another file format depends on the target application. You will find recommended True Page applications in the *Save Files as Type* drop-down list when you save your recognized document. These file formats — marked with a *TP* before their names — are the most likely to successfully retain True Page formatting.

If you use the *Multiple Column* zoning method with True Page, save your document in one of the recommended file formats. All of these formats support frame-based layouts.

Working with Frames in a Target Application

You should be familiar with the way your target application handles frames if you recognize your document with the *Multiple Columns* zoning method and True Page. Different applications may use other terminology to refer to frames such as *boxes* or *objects*. Experiment with your application and consult its documentation to learn the best ways of working with frames.

The following table provides a brief overview of working with True Page documents in some recommended applications.

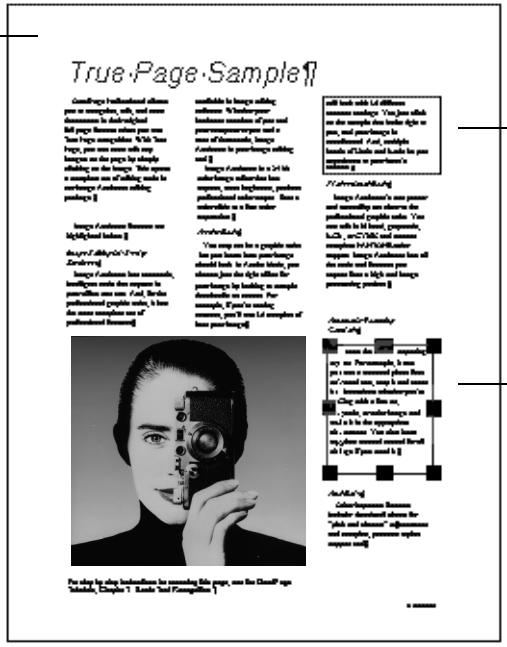
Application	How to View a True Page Document	How an Application Refers to *Frames*	Editing Frames — Getting Started
Microsoft Word for Windows 6.x	Choose *Page Layout* in the View menu.	Frame	Click an area of text to display its frame border. Click the border; handles appear for resizing and moving the frame. To edit text within the frame, place the pointer in the frame and select text to start editing.
WordPerfect for Windows 6.x	Appears in True Page format automatically.	Frame or box	Click an area of text to display its box border. To enter edit mode, place the pointer in the box and double-click. You can edit one text box at a time.
Ami Pro (3.0)	Choose *Layout Mode* in the View menu.	Frame	Double-click a frame to enter edit mode.

Compare your document's format and layout to the original document when you open it in your target application. Edit text and paragraph formatting within each frame, if necessary, before adjusting frames to make layout changes.

Example — Word for Windows 6.0a

The following document was recognized with True Page output and saved in a Word for Windows 6.0a file format.

This page is displayed in *Page Layout* view so that you can see the actual look of it. In *Normal* view, it appears in a one-column layout.



Frames are normally invisible. Click a frame to see its border. To see all frame borders, choose *Options...* in the Tools menu and select the *Text Boundaries* option in the *View* tab.

Click a frame border to select it; handles will appear. You can resize and move the selected frame using its handles.

To add text to a frame, increase the size of the frame. You may need to adjust other frames to make room.

Chapter 7

Improving Performance

This chapter gives specific and general advice on how to speed up recognition, make recognition more accurate, and streamline your OCR workflow.

The following topics are included:

- Improving Speed
- Improving Accuracy
- Legal Documents
- Spreadsheets and Tables
- Foreign-Language and Multilingual Documents
- Scanning Large Jobs

Improving Speed

Although OmniPage is designed to run automatically, the automatic features can take longer to work. You may find you need to trade off better recognition for faster recognition or vice versa. These factors most affect processing speed in OmniPage:

- Computing Power
- Permanent Windows Swap File
- Settings Panel Options
- Manual Zones

Computing Power

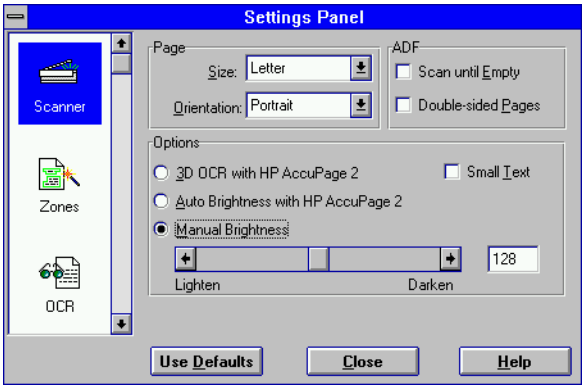
This affects speed the most. A Pentium computer is faster than a 486, which is faster than a 386 and so on. You should have a minimum of 8MB RAM, but as with most CPU-intensive programs, more memory improves processing speed in OmniPage.

Permanent Windows Swap File

To increase speed, set up a permanent Windows swap file (virtual memory) with at least 8MB of free, contiguous disk space. For more information, see “Setting up a Windows Swap File” on page 6.

Settings Panel Options

Click the Settings Panel button in the toolbar or choose *Settings Panel...* in the Settings menu to open the Settings Panel.



Scanner Settings for Faster Recognition

- *Page:* Choose the proper page orientation.
- *Options:* Select *Manual Brightness.*

  Use the *Manual Brightness* control in the Settings Panel Scanner options if you are scanning high-quality printed documents with crisp, black text printed on a white background.

  Although faster, Manual Brightness may not provide the best accuracy for all documents.

See "The Language Analyst and Accuracy" on page 213 for information on how this OCR settings panel option affects accuracy.

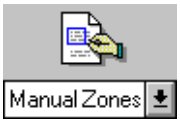
OCR Settings for Faster Recognition

- *Input Options:* deselect *Automatically Correct Page Orientation.*
- Deselect *Use Language Analyst.*

  If you scan high-quality documents with crisp, black letters printed on white paper, recognition is faster with the Language Analyst turned off. This is because the Language Analyst requires time to evaluate words, compute likely errors, and determine replacement words. Turning off the Language Analyst increases recognition time somewhat, but may decrease recognition accuracy, depending on your document.

See "The Language Analyst and Accuracy" on page 213 for information on how this OCR settings panel option affects accuracy.

Manual Zones



Select *Manual Zones* in the drop-down list under the Zone button if you only need a portion of a page recognized or want to ensure that all areas are zoned properly. You can draw selection boxes around just the parts of a page you want recognized. This saves OmniPage the time of recognizing text or graphic elements you do not need.

See "Manual Zones — Recognize Portions of a Page" on page 48 for detailed instructions on how to use the *Manual Zones* feature.

Improving Accuracy

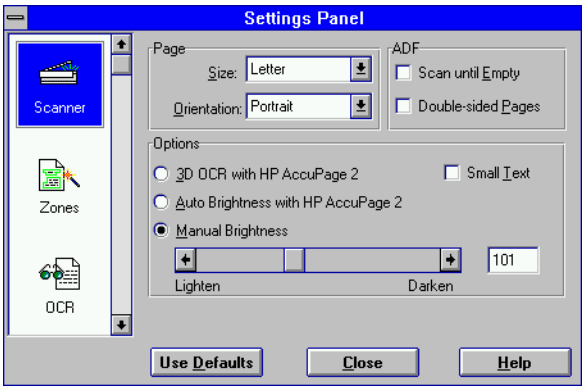
Typeset, high-quality printed pages return the best recognition accuracy.

With lesser-quality pages, text-recognition accuracy will be poorer. These factors most affect text-recognition accuracy:

- Settings Panel Options
- Line Art
- Document Quality
- Scanning Angle
- Scanner Glass Clarity
- Paper Transparency

Settings Panel Options

Click the Settings Panel button in the toolbar or choose *Settings Panel...* in the Settings menu to open the Settings Panel.



Scanner Settings and Accuracy

- *3D OCR with AnyPage/HP AccuPage 2:* Select this setting to scan poor-quality pages, pages with very small type, or pages with text on colored or shaded backgrounds. Used with the Language Analyst, this setting provides the best recognition accuracy possible. It is only available with grayscale scanners.
- *Auto Brightness with AnyPage/HP AccuPage 2:* Select this setting to scan text on colored or shaded backgrounds, or smudged paper. It is slightly less accurate but slightly faster than *3D OCR with AnyPage*. It is only available with grayscale scanners.

- *Manual Brightness:* Select this setting if you are scanning high-quality printed documents with crisp, black text printed on a white background. This is the only option available if you have a black-and-white scanner.

  If text characters on your document appear to be thick and overlapping, adjust the brightness slide towards *Lighten*. If characters appear to be thin and broken, adjust the setting towards *Darken*.

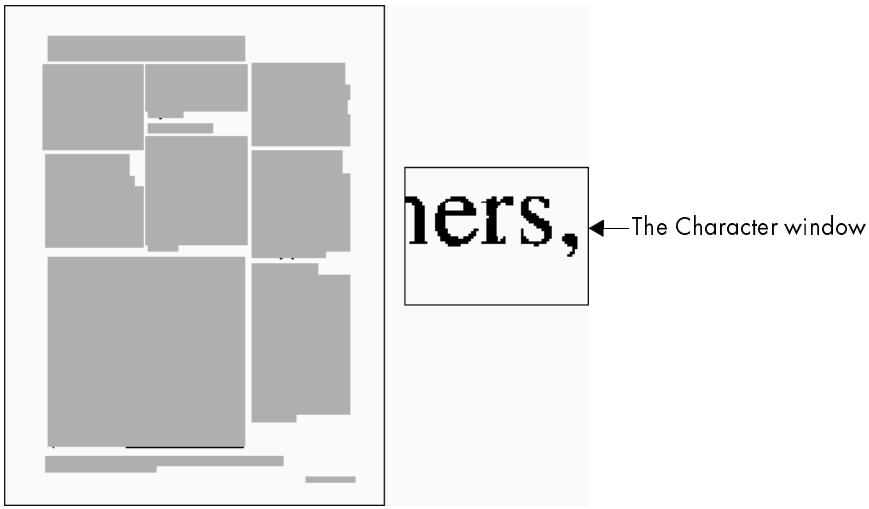
You also can evaluate a document during recognition. See the next section, "The Character Window."

The Character Window

Sometimes a small manual adjustment to brightness can make a big difference in recognition accuracy for users who have black-and-white scanners. Manually adjusting brightness can also be helpful on pages with especially dark shading.

Rescan any document with numerous recognition errors and observe the Character window during OCR.

The Character window appears while OmniPage performs text recognition. It shows samples of the scanned image as OmniPage sees them. Adjust the *Manual Brightness* control in the Scanner settings panel as necessary after you observe characters in the Character window.



This figure shows how well-formed characters appear in the Character window. No special brightness adjustment is needed.

This figure shows how thin, broken characters appear in the Character window. Drag the brightness control toward *Darken* and rescan.

This figure shows how thick, run-together characters appear in the Character window. Drag the brightness control toward *Lighten* and rescan.

The Language Analyst and Accuracy

Select *Use Language Analyst* in the OCR Settings Panel for improved accuracy on most documents. The Language Analyst uses information about language context and usage rules to evaluate characters, compute likely errors, and determine replacement words. It does slow recognition speed somewhat.

Line Art

OmniPage may try to recognize some line-art graphics as text if the artwork is sparse and the lines resemble letter strokes. For accurate recognition of this kind of graphic, draw a manual zone around each graphic and select *Graphics* in the zone contents drop-down list.

Document Quality

OmniPage recognizes characters in almost any font from 6 to 72 points in size. However, keep the following in mind when using OmniPage:

- The print should be reasonably clean and crisp. Characters must be distinct: separated from each other and not blotched together or overlapping.
- The document should be free of notes, lines, or doodles. Anything that is not a printed character slows recognition, and any character distorted by a mark will be unrecognizable.
- The document font should be non-stylized; for example, OmniPage may not recognize the *Zapf Chancery* font accurately.
- It is difficult to recognize underlined text because the underline changes the shape of descenders on the letters q, g, y, p, and j.

Scanning Angle

Make sure that the document is positioned correctly in your scanner and is not slanted. Even if you put a page in the scanner correctly, it is still possible for the page to be turned slightly so that text will be difficult to recognize. The final document may have missing characters, split lines of text, or words recognized incorrectly if the page was not scanned correctly.

If you notice that the page is crooked in the zone window, adjust and rescan it. You may also see slanted text in the Character window. If you are scanning a multiple-page document and notice poor recognition on certain pages, it may be that those pages were crooked in the scanner. Try scanning them again.

Scanner Glass Clarity

The sheet of glass on the flatbed of the scanner must be clean and clear. If it gets dirty, wipe it gently with a soft, damp, lint-free cloth or tissue. Be sure it is completely dry before you put pages on it.

Paper Transparency

Some paper is thin enough that the scanner sees text printed on the opposite side of the scanned page. This is often the case with telephone book pages. To correct this problem, put a black piece of paper behind the page between the page and the lid of the scanner.

Legal Documents

This section lists some tips to keep in mind when scanning legal documents.

General Tips

Keep these general tips in mind when scanning legal documents:

- Select *Legal* size in the Scanner settings panel if the document is printed on legal-size (14 inches in length) paper.
- Many legal documents consist of page-wide text. If this is the case, select *Single Column or Table* as the zoning method in the Settings Panel.
- Experiment find the best process for scanning and saving each document.

Pleading Papers

Keep these tips in mind when scanning pleading papers:

- Generally you should select the *One Zone* zoning method in the Settings Panel.
- Select the *True Page - Retain All Page Formatting* OCR option in the Settings Panel if you would like to reproduce page layout without much editing or reformatting.
- You may want to draw manual zones in some circumstances. If numbers on pleading papers are printed less than five spaces from the text, OmniPage will consider the numbers to be part of the text body and place them on a text line. Try drawing a zone around just the body text to omit the line and line numbers from the recognition process.

  This is a good option if you are going to add text that will change the line numbers. Use your word-processing application to number each line.
- Try scanning or loading your document in Direct Input mode if you want a carriage return inserted at the end of each line. Initiate Direct Input from a word-processing program that is compatible with Direct Input. Select *Use Hard Carriage Returns* in the Direct Input settings panel. Use your word-processing program's commands to number each line. See Chapter 5, Direct Input for information on this option.
- Experiment on a sample page to find the best process for scanning and saving each document, and then try scanning the entire brief.

Spreadsheets and Tables

Keep these general tips in mind when scanning spreadsheets, charts, tables, single-column pages, or memos with page-wide text and tabs:

- Select *Landscape* as the orientation in the Scanner options section of the Settings Panel if the document is presented in landscape view.
- Select *Single Column or Table* as the zoning method in the Settings Panel to preserve the spreadsheet format. When OmniPage detects five or more spaces, the *Single Column or Table* option converts the spaces to a tab.
- Manually zone a table of numbers and select *Numeric* in the *Zone Contents* drop-down list in the zone window to improve recognition. You can create new zone contents files for special characters that your spreadsheet may contain. See “Save Zone Template” on page 100.

Foreign-Language and Multilingual Documents

For the best possible recognition of documents in any language, select:

- The appropriate language(s) in the Select Languages dialog box (choose *Select Languages...* in the Settings menu).

- The main dictionary appropriate for the language you are recognizing in the Spelling section of the Settings Panel.
- The appropriate user dictionary, if you have more than one.
- *Use Language Analyst* in the OCR settings panel.

Foreign-Language Documents

Double-check that your settings are correct as described above when recognizing a foreign-language document. Always select the main dictionary that matches the language you want to recognize. The Language Analyst consults both the main dictionary and the user dictionary during recognition.

OmniPage can still perform OCR on a page if the right dictionary is not available because it recognizes the characters of the language(s) selected in the Select Languages dialog box. Deselect the option *Use Language Analyst* in the OCR settings panel when the right dictionary is not available. Otherwise, OmniPage spends time trying to match words to the wrong dictionary. When it sees that dictionary entries are not improving recognition results, it turns the Language Analyst off anyway.

If you recognize many documents in a language other than that of your default main dictionary, you can order a dictionary for that language from your Caere distributor or by calling Caere in the United States at 800-535-SCAN.

Multilingual Documents

You may want to recognize a document written in more than one language. Select both the proper language set and main dictionary. You can zone multilingual documents automatically or manually. Auto zoning saves processing time but manual zoning may return better recognition results.

Auto Zoning

Suppose you want to recognize a document written mostly in French with a few sections in Portuguese. (Substitute the language(s) you need as necessary in the following instructions.)

1 Choose *Select Languages...* in the Settings menu.

2 Select both *French* and *Portuguese* in the Select Languages dialog box.

3 Click *OK*.

4 Click the Settings Panel button or choose *Settings Panel...* in the Settings menu.

5 Click the Spelling icon.

6 Select your French dictionary in the *Main Dictionary* drop-down list in the Spelling settings panel.

7 Click the OCR icon.

8 Deselect *Use Language Analyst* if it is selected.

9 Set other options as appropriate and click *AUTO*.

10 Click the Check Recognition button to open the Check Recognition dialog box and correct any mistakes when recognition is done.

Manual Zoning

This process is more time-consuming but the results are more accurate.

Recognize the French Portion

1 Select *Manual Zones* in the drop-down list under the Zone button.

2 Follow steps 1 through 6 in the previous "Auto Zoning" section.

3 Click the OCR icon.

4 Select *Use Language Analyst* if it is not selected.

5 Scan or load your document.

6 Draw zones around just the French portions of the text when OmniPage opens the zone window.

7 Click the OCR button when you are done zoning.
OmniPage recognizes the text and opens it in a text window.

8 Click the Check Recognition button to open the Check Recognition dialog box and correct any mistakes.

Recognize the Portuguese Portion

1 Select your Portuguese dictionary in the *Main Dictionary* drop-down list in the Spelling settings panel.

2 Rescan or reload your document.
OmniPage adds a second page to the current Caere Document.

3 Draw manual zones around the Portuguese portions of the text when OmniPage opens the zone window.

4 Click the OCR button when you are done zoning.
OmniPage recognizes the text and opens it in a text window as page two of the current Caere Document.

5 Click the Check Recognition button to open the Check Recognition dialog box and correct any mistakes.

6 Save the document as a word-processing file.

7 Use your word-processing program commands to cut and paste the text as necessary..

Scanning Large Jobs

You can place a large stack of documents in your scanner if you have an automatic document feeder (ADF). Set the appropriate process settings (*Scan Image, Auto Zones* or a saved zone template, *Perform OCR* or *Defer OCR*) and click the *AUTO* button. You may want to fill your ADF and click the *AUTO* button before leaving the office, for example.

OmniPage will process all documents according to your settings. You can then save the documents as a single file or as several smaller files.

You must select either *Auto Zones* or a zone template if you want to automatically process a batch of documents unattended. If *Manual Zones* is selected, OmniPage stops after scanning each page so you can draw zones. If you select *Load Image*, you must select all images to load before processing can continue.

You can save scanned files in word-processing or Caere Document format. Only Caere Document files can be reopened in OmniPage.

To protect your processing time investment, save scanned documents in Caere Document file format before checking recognition or editing them.

Preparing Documents for the ADF

Suppose you want to scan 25 pages. Decide how you will save the scanned documents before you fill the ADF. This affects how you group them in the document feeder.

If you choose a word-processing format, you have three options for saving the scanned pages: as a single file, as one file per page, or as one file for every blank separator that OmniPage locates.

- *Create one file for all pages*

  The pages would be saved as one 25-page file.

- *Create one file per page*

  The pages would be saved as 25 one-page files.

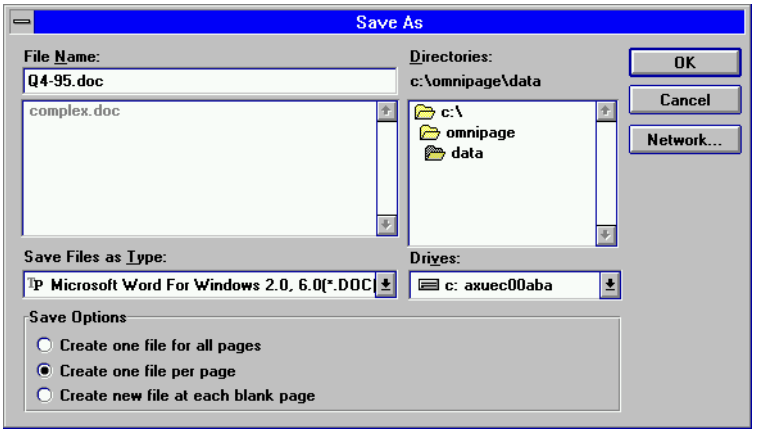
- *Create new file at each blank page*

  Insert blank pages as separators into a stack of one-sided documents. All pages following a blank page would be saved as a separate file with a unique document name.

Saving the File(s)

1 Click the Save As... button in the toolbar when recognition and any text editing you want to do are done.

The Save As dialog box appears.



2 Choose either a word-processing or a Caere Document file format in the *Save Files as Type* drop-down list.

3 Select an option under *Save Options* (for formats other than Caere Document format):

- *Create one file for all pages*: enter a standard eight-character file name.

- *Create one file per page*: enter a name with up to five characters.

  OmniPage adds three numbers to each file name to make it unique. For example, if you typed `file` in the *File Name* box, the first page is saved as `file001`, the second page as `file002`, and so on.

- *Create new file at each blank page*: enter a name with up to five characters.

  OmniPage adds three numbers to each file name to make it unique. For example, if you typed `file` in the *File Name* box, the first batch is saved as `file001`, the second batch as `file002`, and so on.

If you chose the *Multiple Columns* and *Retain Graphics* options in the Settings Panel, the graphics will be saved only if the format in which the file is saved supports embedded graphics. ASCII text, for example, does not support embedded graphics.

You can export graphics separately. See “Export Image” on page 93.

Chapter 8

Technical Information

Use this chapter if you have trouble getting OmniPage or your scanner to run properly, if many errors appear in your recognized text, or if you want to speed up performance. See also Chapter 7, Improving Performance.

The following topics are covered in this chapter:

- Before You Begin
- Installation
- Scanning
- OCR
- Operation
- Supported Output File Formats
- Supported Input File Formats
- Error Messages
- Caere Product Support

Many error messages in OmniPage are self-explanatory and offer a solution. Some are explained in more detail in this chapter; these are listed alphabetically in the Error Messages section starting on page 240.

Before You Begin

If you have a problem with OmniPage, first check that your computer, scanner, and other applications are functioning properly.

System Setup Checklist

- Make sure that your system meets all requirements as listed in “System Requirements” on page 4.
- See that your scanner is plugged in and turned on and that all cable connections are secure before starting OmniPage.

- Use the software that came with your scanner to verify that it works properly before using it with OmniPage.
- Fix problems that occur with Windows before using OmniPage again.
- Run virus-checking software regularly.
- Defragment your hard disk occasionally (see your DOS documentation).

Try this First

It may help to:

- Restart your computer.
- Restart your scanner.
- Reinstall OmniPage Pro.

Always restart Windows if an application hangs or crashes. Try to save open files in all applications before restarting.

Installation

This section contains information on common installation problems as well as information on how to optimize the installation process:

- Installing OmniPage with the Norton Desktop
- Speeding Up Installation
- Conflicts with Disk-Caching Programs
- Using EMM386.EXE
- Setup Program Requests Same Disk

Installing OmniPage with the Norton Desktop

Some versions of the Norton Desktop are incompatible with the OmniPage installation program. If you run Windows under Norton Desktop and you have difficulty installing OmniPage, follow this procedure:

1. Open the Windows `system.ini` file in a text editor and change the line `Shell=ndw.exe` to `Shell=progman.exe`.

   This will make the Windows Program Manager appear when you start Windows rather than the Norton Desktop.

   Consult your Windows User's Guide for more information.
2. Save the file.
3. Restart Windows.

4 Install OmniPage according to the instructions in Chapter 1, Installation.

5 Reopen the Windows `system.ini` file and change the line `Shell=progman.exe` back to `Shell=ndw.exe`.

This will make the Norton Desktop appear when you start Windows.

6 Save the file.

7 Restart Windows.

Speeding Up Installation

OmniPage installation is faster if you use a disk-caching program. Load SMARTDrive, the disk-caching program provided with Windows unless your system uses a different disk-caching program. Refer to the documentation that came with your system.

Conflicts with Disk-Caching Programs

Some disk-caching programs interfere with memory allocation and will prevent you from installing successfully. If you are using a disk cache program other than Windows SMARTDrive, temporarily disable it and try to install OmniPage again. Then enable the cache program and verify that it works.

Using EMM386.EXE

You may not be able to use this memory manager with OmniPage. Try running OmniPage with just the default Windows memory management programs. Consult your memory manager's documentation for instructions on how to deinstall `emm386.exe` from your `config.sys` file.

Setup Program Requests Same Disk

Eject the disk, exit the program and restart your computer if the Setup program repeatedly requests the same installation disk. Try running Setup again.

If this does not solve the problem and the correct disk is in the disk drive, the disk may be damaged. To check the disk:

1 Exit the installation program and Windows.

2 From the DOS prompt, type `dir B:` (if you are installing from drive A, type `dir A:`).

An error message from DOS means that the floppy disk is damaged.

3 Try to copy a file from the OmniPage disk to your hard disk if you do not receive an error message and DOS displays the disk directory.

DOS may be unable to copy files from the disk even if it can read the directory. This indicates disk damage. If the disk is damaged, contact Product Support for a replacement.

Scanning

This section describes common scanning problems and their solutions. Topics include:

- No Scan Image Command
- System Hang During Scan
- Scanner Message on Launch
- Checking the Scanner Driver Name and Version
- System Crash During Scan
- Scanner Hardware Compatibility
- Microtek Scanners
- TWAIN Scanners
- Deinstalling the Scan Manager

No *Scan Image* Command

No *Scan Image* command appears in the Process menu or in the Image button drop-down list if you have not installed the Scan Manager. You must install the Scan Manager if you intend to scan in OmniPage. See “Scan Manager Installation” on page 8 for instructions.

System Hang During Scan

Possible Causes

A system hang during a scan could have several possible causes:

- An interrupt conflict between your scanner and another device such as a bus mouse or network card.

  Check interrupt addresses for a possible conflict.
- A SCSI termination problem.

  Make sure the scanner is terminated properly.
- A memory-related problem (see next section).

Memory-related Problem

Low memory can cause scanning problems. Try closing open windows and applications to free up memory. Try modifying your `system.ini` or your `config.sys` file if freeing up memory does not help.

Modifying the `system.ini` *File*

1. Locate and open your `system.ini` file.
   Its default location is `c:\windows`.
2. Locate the section `[386ENH]`.
3. Add the line `EMMExclude=A000-EFFF`.
4. Below it, add the line `NoEMMDriver=ON`.
5. Choose *Save* and then *Exit* in the File menu.
6. Restart Windows.

Modifying the `config.sys` *File*

1. Locate and open your `config.sys` file.
   Its default location is `c:\`.
2. Locate a line similar to one of the following:
   `device=c:\dos\emm386.sys` or
   `device=c:\qemm\qemm386.sys`.
3. Type `rem` (followed by a space) in front of this line.
   This prevents the expanded memory manager, which can cause conflicts, from loading.
4. Choose *Save* and then *Exit* in the File menu.
5. Restart Windows.

Remove the `rem` from the modified line if this does not help and try this:

1. Locate and open your `config.sys` file.
   Its default location is `c:\`.
2. Look for a line that begins with `devicehigh` or `loadhigh`.
   Some systems may have multiple `loadhigh` lines. Make sure this line is for your scanner driver.
3. Type `rem` (type a space after `rem`) in front of this line.
   This cancels the command line and prevents your scanner driver from being loaded into high memory where OmniPage may not be able to find it.

4 Add a line below it that reads `device=drive:\path\filename` in which drive, path, and filename represent the path to your scanner driver.

This is a typical `config.sys` entry for an HP ScanJet scanner driver:

```
DEVICE=C:\DESKSCAN\SJII.SYS
```

5 Choose *Save* and then *Exit* in the File menu.

Scanner Message on Launch

You may receive the following message in a dialog box the first time you launch OmniPage after installing or changing your default scanner:

"This scanner's configuration is set using the system-level driver."

This is a Windows-generated dialog box. Click *OK* in the dialog box if it asks for no more information. You may have the option of selecting a SCSI ID or scanner configuration information. Consult your scanner documentation for information. You may have the option of entering a page size. Enter the largest size page you may need to scan that your scanner supports and click *OK*.

If you receive a message that says something similar to "Cannot find AccuPage.dll" or "Can't find scanner!" see the message that begins "Problem with scanner" on page 245.

Checking the Scanner Driver Name and Version

The following files allow OmniPage to communicate with your scanner hardware:

- *Scanner driver* files, supplied by the scanner manufacturer, that tell your computer how to communicate with the scanner. You may not need one, or you may need more than one. Refer to your scanner documentation.
- Scan Manager files supplied by Caere that are placed on your hard drive during installation of the Scan Manager application. See "Scan Manager Installation" on page 8.

Scanner Drivers

A scanner driver must be installed on your hard disk. This file tells your computer how to communicate with the scanner. The file name is referenced in your `config.sys` or `autoexec.bat` file.

The file name in the `config.sys` or `autoexec.bat` file must match the file name of the device driver supplied with your scanner software. The file version must be compatible with OmniPage.

This is a typical `config.sys` entry for an HP ScanJet IIp scanner driver:

```
DEVICE=C:\DESKSCAN\SJII.SYS
```

Check your scanner documentation to find the device driver name and version for your scanner. Compare this with the information in the `config.sys` or `autoexec.bat` file. Make sure the two are the same.

For information on how to edit your `config.sys` or `autoexec.bat` file, see your Windows or DOS documentation. Restart your computer when you finish editing so that the new device driver loads.

Record any parameter information for the device driver such as port address and speed. This information may be required when you select your scanner for the first time or if you change your scanner installation. See your scanner documentation for information on these settings.

Load only one device driver for your scanner at a time. Multiple device drivers for the same scanner may cause problems.

Extended Memory Manager

If you use an extended memory manager, avoid loading your scanner driver high in memory. Consult your memory manager documentation for detailed information about loading drivers high.

System Crash During Scan

If you experience a system crash when you try to scan, try modifying your `system.ini` file. Follow steps 1 through 3, and then step 5 in the section "Modifying the system.ini File" on page 226. Restart Windows.

Scanner Hardware Compatibility

If you experience a problem between OmniPage and your scanner, make sure the hardware used matches the hardware OmniPage supports. Check the Supported Scanners list in the Release Notes.

For some scanners, you will also need to know the port address and speed and enter them in the *Hardware Options* section of the Scan Manager dialog box. Refer to you scanner documentation for more detailed information.

Microtek Scanners

Set scanner speed to `2` for best accuracy. Consult your Microtek scanner documentation for instructions on how to set scanner speed.

TWAIN Scanners

To select a supported TWAIN scanner:

1. Exit from OmniPage if it is open.
2. Open the Scan Manager in the Caere Applications program group if you have installed the scanner application already.
   Otherwise, install the Scan Manager according to the instructions in Chapter 1.
3. Click *Add>>* to expand the dialog box.
4. Insert the Scan Manager disk in your floppy drive when instructed in order to see a list of available scanner drivers.
5. Select *TWAIN scanner* in the *List of Scanners* list box and click *Install.*
   The Select Source dialog box appears.
6. Select a TWAIN source and click *Select* to return to the Scan Manager dialog box.
7. Click *Set As Default Scanner.*
8. Click *Close.*

Do not run other TWAIN-compliant applications while using or installing a TWAIN scanner with OmniPage.

Use Relisys and Umax Scanners with TWAIN Only

OmniPage supports Relisys scanners and certain Umax scanners only with the TWAIN interface. Select *TWAIN scanner* in the Scan Manager Installation dialog box when you install a Relisys driver or a Umax UG80, UC630, or later driver.

Scanning with the Canon CJ-10 Scanner

This scanner requires specific settings to work properly with OmniPage.

Canon CJ-10 users can only select *Manual Brightness* in the Scanner settings panel. The other settings are disabled.

A Canon dialog box appears after choosing *Scan Image* in the Process menu or clicking the Image process button in the toolbar. Unlike other TWAIN scanners, the CJ-10 requires an 8-bit-wide scan with the manual brightness setting instead of a 1-bit-wide scan. Select the following settings:

- 256-level grayscale
- 300 dpi resolution

Using Microtek Scanners with TWAIN

The *Manual Brightness* setting in Scanner settings panel is not adjustable when using the TWAIN interface with Microtek scanners.

Deinstalling the Scan Manager

You may want to deinstall the Scan Manager application if you are not using a scanner with OmniPage. This frees up some disk space and prevents an error message saying that OmniPage cannot find the scanner.

Delete the following files in your `windows` directory:

`scnlib.dll, sccpl.dll, accupage.dll, apsl.exe, aplib1.dll, aplib2.dll, sg5w30.dll, hpscaner.dll`

Delete the following files in your `windows\system` directory:

`pizperm.dll, pixdflt.dll, pixpnr.dll`

Select the Scan Manager icon in the Caere Applications program group and press the Delete key on your keyboard to delete it. Click *Yes* in the dialog box that asks if you are sure.

OCR

This section describes common OCR problems and their solutions:

- System Hang During OCR
- Text Not Recognized in Manual Zones
- Document Image Problem in Zone Window
- Fax Recognition Problems
- Formatting Issues
- OmniPage OCR Tests

Text Not Recognized in Manual Zones

This can happen when you have selected zones with very small text and OmniPage mistakes the text for a graphic. Try the following:

1 Open the Settings Panel.

2 Click the Zones icon.

3 Select *One Zone* as the zoning method.
OmniPage will try to recognize all selected zones contents as text.

4 Re-recognize the document.

Document Image Problem in Zone Window

You may see vertical lines running through the document image or no image in the zone window and then receive garbage or nothing after OCR.

The memory address for your scanner interface card is probably interfering with the memory address for your video display adaptor. Use the instructions in your scanner documentation to move the scanner interface card to a different memory address.

You could also try using a standard VGA video driver.

Fax Recognition Problems

Recognizing faxes accurately can be difficult. A typical fax machine produces documents at 200x100 or 200x200 dpi. A typical scanner scans at 300x300 dpi. This means there is less information to work with on a fax. Recognition accuracy may not be as good as on the printed original. Also, faxes printed on thermal paper are more difficult to scan than a document on regular white paper. Try this to improve fax recognition accuracy:

- Have senders select Fine (or Best) Mode when they send you a fax. This sets scan resolution at 200x200 dpi.
- Have senders transmit files directly to your computer via fax modem if you both have one.

  Fax modem transmissions are much clearer than fax machine transmission because there is no scanning involved. Load the fax file as an image in OmniPage and then recognize the text.
- It also helps if senders use a sans-serif font that is 11 points or larger on clean white paper. There should be no marks, smears, or doodles on the page.

Formatting Issues

This section provides workarounds for possible formatting problems.

True Page and WordPerfect for Windows 5.2

You cannot scroll a one-page document containing True Page output in WordPerfect for Windows 5.2. To view a document, try recognizing your document with the *Retain Font and Paragraph Formatting* option in the OCR settings panel instead of *True Page - Retain All Page Formatting* .

True Page and WordPerfect for Windows 6.0

You can save a maximum of 12 pages of True Page output to WordPerfect for Windows 6.0. If your document exceeds 12 pages, you can do one of the following:

- Choose *Delete Current Page* in the Edit menu to delete excess pages.
- Select *Create One File per Page* under *Save Options* in the Save As dialog box.

The first frame border in a True Page document may have its position and size distorted when you open it in WordPerfect. You can edit this frame in WordPerfect. This problem does not occur with the WordPerfect 6.0a upgrade; please contact Novell/WordPerfect for upgrade information.

OCR and Wide Text Zones

Portrait pages recognized in OmniPage should not have a single text zone wider than 7.5 inches or the right margin may be incorrect in your target application. If your page has a multicolumn layout wider than 7.5 inches across, separate it into zones of fewer multiple columns. If the page has one zone wider than 7.5 inches, recognize it as a single zone. Adjust the right margin in your target application as necessary.

Spreadsheet Applications

Spreadsheet-type documents may require a special OCR procedure if you plan to save to a spreadsheet application such as Excel. If your document has separate areas of information with different row and column alignment, recognize each area individually and save to separate files. You can later combine these files in your spreadsheet application.

For example, if you have a spreadsheet with two separate text areas divided by a heading, follow this procedure:

1. Draw a single zone around the first text area.

   The information within the zone should all be in the same row and column alignment. Do not include headings, titles, footers, and so forth.

2. Select *Single Column or Table* in the Zones settings panel and *Retain Font and Paragraph Formatting* in the OCR settings panel.

3. Perform OCR and then use the *Save As...* command to save the recognized text to the desired spreadsheet file format.

4. Return to the zone window and delete the manual zone.

5. Repeat steps 1–3 for the other area of data in the page image that you want to recognize.

6 Save the recognized text in a separate file.

7 Combine the resulting spreadsheet files in your spreadsheet application using cut and paste commands.

Start with the file that has the cell format that you want. Cut and pasted material should conform to this format.

Incorrect Font Size Output

The TrueType font size output in your recognized document may be incorrect if your screen resolution is set to large-size fonts. To test the font size output, perform OCR on the `test.tif` image file in OmniPage. The font size of the recognized text should be 11 points. If the font size is larger than this, try setting a different system font size.

1 Open Windows Setup in the Program Manager Main window.

2 Choose *Change System Settings...* in the Options menu.

3 Consult your Windows or PC documentation to decide which setting you can use.

Perform OCR with different settings until the font size is correct.

Retaining Graphics With *AUTO* and HP AccuPage 2

If you use an HP Scanner that supports HP AccuPage technology, select *3D OCR with HP AccuPage 2* or *Auto Brightness with HP AccuPage 2* in the Scanner settings panel to retain graphics when you use the *AUTO* button to process documents. Do not select *Manual Brightness* or graphics will be ignored.

System Hang During OCR

Many computer systems provide a feature called shadow RAM to enhance system performance. If OmniPage causes the system to hang, turn off the shadow RAM function of your computer and try again. Refer to the computer reference manual for information on disabling shadow RAM.

Some computer systems do not allow you to turn shadow RAM off. Incompatibilities with these systems are usually not related to shadow RAM.

System hangs may be related to incompatibilities with memory-resident applications or device drivers. Use a text editor to comment out any memory-resident device drivers and applications from your `autoexec.bat` and `config.sys` files not used by Windows, OmniPage, your scanner, or your hard drive and reboot your system.

Do not remove a device driver unless you are aware of its function and know it may be removed safely.

Hard disks often require special device drivers that should not be removed. Video displays that require special device drivers may need to be reconfigured instead of removed. Make a backup boot disk with your current operating system version, `autoexec.bat`, and `config.sys` to guard against potential mistakes.

OmniPage OCR Tests

If OmniPage seems to run slowly even with all other applications closed, load the `test.tif` file or scan one of the sample pages included with your OmniPage package to find the typical recognition time for a good-quality document and image on your system.

Successfully completing the test procedure does the following:

- Provides a benchmark for recognition time on your system.
- Verifies recognition accuracy independently of your document or image quality.

The test file should produce a text file with near-100% text recognition accuracy. If you are unable to achieve a similar level of accuracy with your documents, review Chapter 7, Improving Performance, for possible problem areas and solutions.

Operation

This section provides some information and troubleshooting techniques for the following topics:

- Low Memory
- Low Disk Space
- Slow Operation
- OmniPage No Longer Works
- Deleting *.set, *.trn, *.ud, *.zcn, and *.zon Files
- Testing OmniPage with a Simplified System
- Required HP Printer Driver
- Orchid Fahrenheit 1280 Video Card

Low Memory

OmniPage runs poorly in a low memory situation. Optimize your system to reduce the possibility. Here are a few tips:

- Increase the amount of physical RAM.

  More memory optimizes OCR performance: 8MB is the minimum required, but 12MB is recommended.

- Use a program for disk caching such as Windows SMARTDrive.
- Use a known compatible extended memory manager such as `himem.sys`.
- Add the line `EMMExclude=A000-EFFF` under the `[386enh]` section of the `system.ini` file. (The characters after the `A` are zeroes.)
- If using Quarterdeck's QEMM386 as the memory manager, locate the `qemm386.sys` statement in the `config.sys` file. Add a NOEMS switch to the line. Remove the stealth feature from this command (`ST:x`). The command should be as simple as possible: `device=\drive\QEMM\QEMM386.SYS RAM NOEMS`
- Defragment your hard disk occasionally to free up contiguous blocks of memory. Type `defrag` at the DOS prompt or check your DOS or Windows documentation.

Error Messages or Slow Operation

You may receive out-of-memory error messages or find that OmniPage works slowly and accesses the hard drive a lot. Try closing open OmniPage windows and any other open applications to free up memory.

Windows Swap File

OmniPage uses the Windows swap file when memory runs low. See "Setting up a Windows Swap File" on page 6 for information. Windows uses disk space to simulate the RAM it does not have available. Your system runs more slowly when it has to use a swap file.

Try closing any other open applications and restarting Windows. This can free enough memory for OmniPage to operate without using the swap file. If you often work with long, complex documents or need to have many documents open at a time, add more RAM to your system.

Amount of Memory Available

Verify that there is at least 8MB RAM available in Windows. Check free memory available in Windows by choosing *About Program Manager...* in the Program Manager's Help menu. You can find out how much memory

you have when you start up your computer. Type `mem` at the prompt. DOS displays your total memory and other information.

For information on optimizing your system and application performance, see your Windows User's Guide.

Low Disk Space

Check your `temp` directory for unnecessary files if you seem to be running out of disk space too quickly. Files are stored here especially after a system hang or crash. Usually the path to this directory is `c:/temp`. Type `set` at the prompt in MS-DOS in order to find your `temp` directory if the path is not in `c:/temp`.

Try to keep at least 20MB of free disk space available on the drive where the temp files are stored.

You also can delete any conversion filter files you do not use. See "Supported Output File Formats" on page 238 for a list of output file formats and the conversion filter file associated with each one. The default location of the conversion filter files is `c:\omnipro`.

Slow Operation

OCR can be a very time- and memory-intensive process, so the processing power of your computer will determine the speed of processes. For example, you will notice a significant speed improvement if you upgrade from a 386 to a 486 or a Pentium.

OmniPage No Longer Works

If you have used OmniPage without difficulty in the past, you may have altered your system configuration. Make sure that your scanner still works with other software. If you have recently installed a new application, this may be the cause. You may have to reinstall OmniPage. Be sure to save your user dictionary (`user.ud`) before doing so.

Deleting *.set, *.trn, *.ud, *.zcn, and *.zon Files

You can delete these files from your hard drive to free up disk space or if you no longer need them:

- Settings files (*.set)
- OCR training files (*.trn)
- User dictionary files (*.ud)

- Zone contents files (*.zcn)
- Zone template files (*.zon)

The default location for all these files is the `omnipro\data` directory. The training files, user dictionary files, and zone contents files must be located here, but a settings file or a zone template file can be saved anywhere on your hard drive. Exit OmniPage and use File Manager or DOS commands to locate and delete unwanted files.

Testing OmniPage with a Simplified System

If OmniPage does not run correctly, try running it on a simplified setup: a system with no network commands, other drivers, memory managers, and so on. This eliminates the possibility of conflicts with any other devices or drivers.

Use a text editor to disable memory-resident device drivers and applications from your `autoexec.bat` and `config.sys` files not used by Windows, OmniPage, your scanner, your hard drive, or your monitor. Restart your computer. See your DOS documentation for instructions on how to edit your `autoexec.bat` and `config.sys` files. Be sure to back up any system files before making changes.

If OmniPage runs on this simplified system, it shows that the problem is a conflict with an item in your system's `config.sys` or `autoexec.bat` files. Add items back to the `config.sys` and `autoexec.bat` files one by one, restart the computer, and start OmniPage. When OmniPage no longer runs, you will know that the last item you added is incompatible with OmniPage.

Required HP Printer Driver

OmniPage requires the `HPPCL5MS` driver for HP printers. You may receive a GPF in the `GDI.EXE` module if you try to print from OmniPage using another HP printer driver. Contact Hewlett-Packard for another driver.

Orchid Fahrenheit 1280 Video Card

Set your monitor resolution to `800x600` in the Orchid Setup dialog box (in the Windows Control Panel). Otherwise, unprintable characters may appear in the text window.

Supported Output File Formats

OmniPage can save files in the following file formats.

Save a file in Caere Document format (*.met) if you want to be able to reopen and work with it in OmniPage.

Application	Conversion Filter File
Caere (*.met)	CAERE01
MasterW4W	CAERE02
Ami Professional 1.1,1.2	W4W33T.DLL
Ami Professional 2.0	W4W33T.DLL
Ami Professional 3.0	W4W33T.DLL
ANSI[1]	W4W01T.DLL
ANSI Standard[2]	W4W01T.DLL
ANSI Stripped[2]	W4W01T.DLL
ASCII[1]	W4W01T.DLL
ASCII Standard[2]	W4W01T.DLL
ASCII Stripped[2]	W4W01T.DLL
dBase IV 1.0	W4W770T.DLL
dBase III, III+	W4W770T.DLL
DisplayWrite (DCA/RFT)	W4W15T.DLL
EBCDIC	W4W02T.DLL
Excel 3.0, 4.0, 5.0[3]	W4W21T.DLL
Excel Text	W4W21T.DLL
Lotus *.WK1	W4W20T.DLL
Microsoft RTF	W4W19T.DLL
Microsoft Word 3.0,3.1	W4W05T.DLL
Microsoft Word 4.0	W4W05T.DLL
Microsoft Word 5.0,5.5	W4W05T.DLL
Microsoft Word For Windows 1.0	W4W44T.DLL
Microsoft Word For Windows 2.0, 6.0	W4W44T.DLL
MultiMate 3.3	W4W10T.DLL
MultiMate Adv. 3.6	W4W10T.DLL
MultiMate Adv. II 3.7	W4W10T.DLL
OfficeWriter 4.0	W4W16T.DLL
OfficeWriter 5.0	W4W16T.DLL
OfficeWriter 6.0,6.1	W4W16T.DLL
PFS:First Choice 1.0	W4W08T.DLL
PFS:First Choice 2.0	W4W08T.DLL
PFS:Prof. Write 1.0	W4W08T.DLL
PFS:Prof. Write 2.0,2.2	W4W08T.DLL
PFS:WRITE Ver C	W4W08T.DLL
Quattro Pro 4.0 (Excel 4.0)	W4W21T.DLL
Quattro Pro for Windows 4.0	W4W21T.DLL
Signature	W4W17T.DLL

Ventura Publisher (MS Word)	W4W44T.DLL
Windows Write 3.0	W4W43T.DLL
WordPerfect 5.0	W4W07T.DLL
WordPerfect 5.1	W4W07T.DLL
WordPerfect 6.0	W4W48T.DLL
WordStar for Windows 1.x, 2.0	W4W04T.DLL
Xywrite III Plus, IV	W4W17T.DLL

[1] This ASCII or ANSI format inserts hard returns to preserve the original line and paragraph breaks and attempts to reproduce white space, including margins, indents, and blank lines. ANSI format is the same as ASCII format with additional support of foreign characters for Windows applications.

[2] Standard ASCII or ANSI format inserts hard returns to preserve the original line and paragraph breaks; use this format for data destined for a system that only recognizes the standard ASCII or ANSI format. Stripped ASCII or ANSI inserts hard returns to preserve paragraph breaks only.

[3] Files will be saved to Excel 2.1 format. This file format can be opened in Excel versions 3.0, 4.0, and 5.0.

Image File Formats

BMP

PCX

TIFF Uncompressed

TIFF Compressed (Types II and IV)

Supported Input File Formats

OmniPage can open or load files with the following file formats:

- BitFax (`*.bfx`)
- BMP
- DCX
- GIF
- JPEG (`*.jpg`)
- MET (Caere Document format; version 5.0 or later)

  Caere Documents cannot be loaded. You must choose *Open...* in the File menu to open a saved Caere Document.
- PCX
- TIFF Uncompressed
- TIFF Compressed (Types II and IV)
- WinFax (`*.fxr`)
- WinFax Lite (`*.fxs`)

- Fax File Formats

  OmniPage supports fax files saved in to DCX, PCX or TIFF format. Many fax boards can receive or convert files in these three formats. Refer to your fax documentation for more information.

TIFF files must be 200, 300, or 400 dpi. 300 dpi is recommended. TIFFs are stored and displayed as 300 dpi line art.

Error Messages

This section contains a list of some error messages that may appear in OmniPage. They are listed alphabetically.

*n*K bytes needed, *n*K available for operation. Please make available some more disk space and try again.

This message means that the operation you are attempting needs more than the maximum amount of virtual memory currently set.

To increase the amount of virtual memory available:

1. Choose *Memory...* in the Setup menu.
2. In the *Temporary file* section, locate the *Maximum* text box.
3. Delete the number in the box and type in a number that is equal to or greater than the number specified in the error message.
4. Click *OK*.

Cannot read file filename.ext.

The Setup installation program presents this error message when it cannot read a file on the OmniPage disk set. See “Setup Program Requests Same Disk” on page 224.

Choose Open Document in the File menu to open Caere Documents.

You tried to open a Caere Document with the *Load Image* command. Open a Caere Document by choosing *Open Document...* in the File menu.

Error accessing the user dictionary. Try freeing up hard disk space. Dictionary size is limited to 32K.

This error message can appear if you are low on disk space. Try freeing up hard disk space. The message can also appear if you have moved

OmniPage to a different location on your hard disk or renamed directories in the path where OmniPage is located. If this is not the case, reinstall OmniPage.

Error adding word to the user dictionary. Try freeing up hard disk space. Dictionary size is limited to 32K.

You are low on free disk space or the user dictionary is full. The user dictionary capacity is 32K or about 5300 words. Free up some disk space or edit the user dictionary to remove unnecessary words. You also can create your own user dictionary with a different name. See "Edit User Dictionary" on page 149.

Error creating a mail document. Be sure your mail application is working correctly. Try freeing up hard disk space in your TEMP directory.

Test your mail application to be sure it is working correctly. Be sure there is at least 1MB of free disk space in your `temp` directory. The default location of the `temp` file is `c:\temp`.

Error creating the zone window. Try closing open windows and applications to free up memory.

See "Low Memory" on page 235.

Error deleting page. Try closing open windows and applications to free up memory.

See "Low Memory" on page 235.

Error displaying the text window. Try closing open windows and applications to free up memory. You may need to exit OmniPage and restart Windows.

See "Low Memory" on page 235.

Error during OCR. Try freeing up hard disk space. Close open windows and applications to free up memory. See the Technical Information chapter in the Reference manual if the problem persists.

This message can appear for a number of reasons: low disk space, low memory, a complex page, or a corrupted file.

Delete unnecessary files to free up disk space. See "Low Memory" on page 235.

The page may have a very complex layout or has very small text and requires too much memory to recognize. Select *Manual Zones* in the drop-

down list under the Zone process button and try drawing fewer, smaller areas on the page.

Sometimes restarting Windows and relaunching OmniPage can clear up problems. If the problem persists, try reinstalling OmniPage.

Error finding zones on the page. The page may be too complex. Try using Manual Zones to recognize smaller areas of the page.

Your page has a very complex layout or has very small text and requires too much memory to recognize. Select *Manual Zones* in the drop-down list under the Zone process button and try drawing fewer, smaller areas on the page.

Error getting the image from the scanner. Exit OmniPage and reselect your scanner in the Scan Manager. Try restarting Windows and your scanner.

There may be a scanner communication problem. Reselect your scanner in the Scan Manager program in the Caere Applications program group. Restart both Windows and your scanner to defragment memory and reset the scanner to its default state.

Error initializing OCR. Try closing open windows and applications to free up memory.

See “Low Memory” on page 235.

Error loading file. The file may be corrupt. Check that the file type appears in the List Files of Type drop-down list.

The file you are trying to load may be corrupt. If so, OmniPage will not be able to use it. Verify also that it is a file type that OmniPage recognizes. See “Supported Input File Formats” on page 239. Reinstall OmniPage if this message continues to appear with other files you try to load.

Error loading the OCR training file. The file may be corrupt or moved. Try reinstalling OmniPage.

You may have moved OmniPage to a different location on your hard disk, or you may have renamed directories in the path where OmniPage is located. Reinstall OmniPage.

Error loading the program. Try closing open windows and applications to free up memory.

An internal program file may have been damaged or is no longer in the OmniPage directory. Try reinstalling OmniPage. If this does not work, call Product Support.

Error loading the training module. The file may be corrupt or moved. Try reinstalling OmniPage.

An internal program file may have been damaged or is no longer in the OmniPage directory. Either restore the file (`rtrain.dll`) or reinstall OmniPage.

Error opening or loading file. Try closing open windows and applications to free up memory. Verify that you are opening a valid file type.

See "Low Memory" on page 235. The file must be in a format that OmniPage recognizes. See "Supported Input File Formats" on page 239.

Error opening the OMNIJOB.INI deferred file list. The file may be corrupt or moved. Try deleting the file and reentering the files for deferred processing.

The `omnijob.ini` file contains the list of documents you have set for deferred OCR. For some reason, OmniPage cannot read the file. It may have been moved elsewhere on your hard drive or become corrupted.

Exit OmniPage and look in your `omnipro\data` directory for the file. If the file is there, it is probably corrupt. Delete it. OmniPage will recreate the file the next time you relaunch it. Reenter deferred files in the *Files to Finish* list box if necessary.

Error opening/removing word from/retrieving word from/the user dictionary. The file may be corrupt or moved. Try reinstalling OmniPage.

You may have moved OmniPage to a different location on your hard disk, or you may have renamed the directories where OmniPage is located.

If this is not the case, exit OmniPage and look in your `omnipro\data` directory for the file `user.ud`. If the file is there, it is may be corrupt. Delete or remove it. (You will lose all data in the file if you delete it; however, the data is lost anyhow if the file is corrupt.) OmniPage will recreate the file the next time you relaunch it. Reinstall OmniPage if there is still a problem.

Error saving file. Try selecting a different file format in the List Files of Type drop-down list.

There is a problem with the document you are trying to save. Try a different file format. If the problem happens consistently with just one file, reload or rescan the image and try to save it again. If this problem happens with different files, try reinstalling OmniPage.

Error saving file to disk. Try freeing up hard disk space. Close open windows and applications to free up memory.

You may be out of disk space or low on memory. Delete unnecessary files to free up disk space. See "Low Memory" on page 235.

If freeing up disk space and memory does not work, the file you are saving may be corrupt. Try rescanning the document or reloading the file.

Error writing to OCR. Rescan or reload the image. You may need to exit OmniPage and restart Windows.

An internal program file may not work with a particular file or may be corrupted. If the problem happens with different files, try reinstalling OmniPage.

For these save options the filename must be less than *n* characters long.

See "Save Options" on page 92 for an explanation file names allowed for various save options.

Insufficient disk space on destination drive. Please free up some disk space and try again.

You have run out of free disk space on your hard disk. Delete any unnecessary files to free up space. See "Low Disk Space" on page 236.

Invalid image file format for this page. Tiff type II and IV support only 1-bit images.

You tried to save an image to a file format that is not supported by OmniPage. Try saving it to another format. See "Supported Output File Formats" on page 238.

OmniPage did not detect any TrueType fonts. Font and formatting information will not be used.

Font mapping only works with True Page when TrueType fonts are installed on your system. Usually, several TrueType fonts are installed when Windows is installed. Refer to your Windows documentation for information on reinstalling the fonts if necessary. Otherwise, OmniPage will display text in a standard non-TrueType font.

Out of memory. Try closing open windows and applications to free up memory.

See "Low Memory" on page 235.

Problem with scanner. Make sure the Scan Manaaer was installed. Check your scanner connections, make sure the scanner is not in use by another application, and verify that you selected the correct default scanner.

Try the solutions listed in the message first. Make sure the scanner is turned on and connected securely. Check your default scanner selection in the Scan Manager. If this does not solve the problem, files may be missing or corrupt. Try reinstalling the Scan Manager.

The page may be too complex. Try creating manual zones and recognizing one half of the page at a time.

The page has a very complex layout or has very small text and requires too much memory to recognize. Select *Manual Zones* in the drop-down list under the Zone process button and try recognizing one half of the page. Save the recognized text, or rescan or reload the image so that it becomes page two of the document. Zone and recognize the second half of the page. Cut and paste the recognized text either in your target application or in OmniPage.

The word already exists in the dictionary.

The text you are trying to enter into the dictionary is already in the dictionary.

There are already 100 files in the deferred processing list. You cannot add any more files.

This message appears when there are 100 files in the *Files to Finish* list box in the Finish Deferred Documents dialog box and you click *Add* to try to add more files. Either process the files already in the list or remove some to make room for new files.

There are no images to save in this document.

This command appears in two situations. It appears if you select *Export Image...* in the File menu but:

- You deselected the *Save Page Images in Caere Document* option in the Preferences settings panel before scanning the document or loading the file.
- You selected *Save Each Graphic Zone to a File* in the *Image Option*s section of the Export Image dialog box but have no graphic zones defined on the page image.

In the first case, open the Settings Panel to the Preferences options and select *Save Page Images in Caere Document*. Then rescan or reload the image.

In the second case, either select *Save Entire Page to a File*, or create manual zones on the page image and specify them as graphic zones and select the *Save Each Graphic Zone to a File* option in the Save As dialog box.

There is no recognized text to save in this document.

You tried to save a document to a text file format before performing OCR. There is no recognized text in a document to save before OCR.

Either recognize your file and then save it or choose *Export Image...* in the File menu to save your unrecognized image as an graphic-format file.

This directory contains more than 100 files. OmniPage can only process 100 files at a time.

This message appears if you select an input directory that contains more than 100 files. OmniPage cannot process more than 100 deferred files at a time. Either remove some files from that directory or select another input directory.

This file could not be processed. Would you like to delete the file? Click Yes to delete the original file permanently. Click No to remove the file from the Files to Finish list box but not delete it.

This message appears if you select a file that has the words *Job Failed* under *File Format* in the Finish Deferred Document dialog box, and then click *Remove*. Because the file may be corrupt, OmniPage gives you the option of deleting the file permanently. Click *Yes* to delete the file permanently. Click *No* to remove it from the *Files to Finish* list box but not delete it from your hard drive. You can also try clicking *Save As*, saving the file to another format, and processing it again. If this does not work, remove it from the Finish Deferred Document dialog box.

This file has not yet been saved.

You chose the *Revert to Saved* command when there is not a saved file to revert to. You must either save the document with the changes you have made or reprocess the document.

This file has the same file name as one already in the list. Change the file name before adding it for deferred processing.

Files in the *Files to Finish* list box in the Finish Deferred Documents dialog box cannot have the same file names even if they have different extensions (such as `file.doc` and `file.met`). Use DOS or File Manager commands to change the file name before adding it to the list.

This is not a recoanized imaae file format. Check that the file type appears in the List Files of Type drop-down list.

You tried to load an unsupported image file.The file type should appear in the *List Files of Type* drop-down list in the Load Image and Open Document dialog boxes when you load or open an image. See "Supported Input File Formats" on page 239.

This is not a valid file name.

The file name must be eight characters or fewer in most cases. See your DOS documentation for a complete list of file naming rules. In general, just use letters and numbers in the file name.

This word cannot be added to the user dictionary because it contains an invalid character (^ or *).

You cannot add words that contain a caret (^) or an asterisk (*) to your user dictionary. You can add words that contain the pound symbol (#) but OmniPage removes the symbol before loading the word into the dictionary. The word `#five`, for example, would be added to the user dictionary simply as `five`.

You can only select Caere Documents or supported image file formats for deferred processing.

You tried to select a file that was not a Caere Document or an image file for deferred processing. See "Supported Input File Formats" on page 239. You can select any files on this list for deferred processing.

You can only select Caere Documents or supported image file formats when you use the Open Document command.

You tried to open a text file that was not a Caere Document file (with the extension `*.met`) or a supported image file format. See "Supported Input File Formats" on page 239.

Caere Document files can only be created by OmniPage.

The following are all incompatible image file formats:

- Fax files that have a proprietary file format
- TIFF files that use a compression scheme or TIFF tags unsupported by OmniPage
- Files not on the Supported Input File Formats lists

Caere Product Support

Product support is available if you need help. However, please check the index or table of contents *first* to find the information you need in this manual — you may be able to save yourself a phone call.

Dial-up Services

Product support and information are available through the following services.‘

Service	How to Contact	Service Provided
Web Site	**www.caere.com**	product information, software updates
America Online	Keyword CAERE	forum, updates
Fax on demand	408 354-8471	common Q&A
Caere BBS	408-395-1631	text, help files, updates
E-mail	support@caere.com	e-mail help

Registered users in North America who need further assistance can also call Caere Product Support at 408-395-8319.

You must be registered in order to receive Product Support! Follow instructions in “Registering OmniPage” on page 12.

Information We Need From You

For the most efficient response, please have the following information on hand and be near your computer when you call:

- **OmniPage serial number *and* version**
- **DOS operating system version**
- **Scanner manufacturer and model**
- **The amount of RAM in your system**
- **The amount of free memory available**
- **A listing of the `autoexec.bat` file**
- **A listing of the `config.sys` file**
- **The size and location of your Windows swap file**
- **The make and model of your computer and peripheral devices**

Chapter 9

Understanding OCR

Caere's OmniPage products represent the leading edge in page- and text-recognition technology. OmniPage Professional and OmniPage can recognize virtually any scanned page, separate text from graphics, and convert almost any printed material to text files for your favorite word-processor, spreadsheet, or database applications.

OmniPage Professional's speed and accuracy are at the forefront of current technology. The integration of *3D* OCR, the Language Analyst, True Page output formatting, training, and 24-bit color image-editing make OmniPage Professional's power and productivity unbeatable.

How OCR Works

OCR is optical character recognition: the process of transferring text from printed pages into an editable computer file — without retyping.

A scanner is more than a copy machine, simply transferring an image into your computer. A scanner translates a page into data by dividing the scanned image into millions of dots or bits (usually from 40,000 to 90,000 per square inch). It then assigns a value to each dot, depending upon whether it is inked, partially inked, or blank.

The composite document stored in your computer is the map of these dots, or a *bitmap*. Your computer sees this data not as editable text, but as one bitmapped image, editable only with image editing tools.

OCR is the process of translating this image into editable text. Text characters are designed by assigning a code corresponding to the keys on the keyboard to each letter, number, or symbol. There are a variety of different code sets in use, but the most common code set is the ASCII (American Standard Code for Information Interchange) table of character equivalents. ASCII is generally recognized as the universal code for small computers. Almost every program that makes use of text and / or numbers understands ASCII.

Prior to 1988, text recognition consisted of a method called matrix-matching. The program compared a bitmap's shape to a library of character shapes. A letter was identified when the shapes matched.

Matrix matching, however, only worked for a small number of fonts and sizes. Because of its limitations, OCR was neither widely used nor recognized as a useful process. This changed when Caere released OmniPage, a page-recognition program that incorporated feature-analysis OCR technology. Feature-analysis OCR analyzes individual character features rather than trying to match shapes as the matrix-matching method did.

Basic OmniPage OCR Technologies

This section details the basic OCR technologies that are used in OmniPage and OmniPage Professional.

When converting the bitmapped image into usable text, OCR software must do the following:

- Analyze the structure of the text on the page to detect the presence of columns and headlines.
- Separate text from graphics on the page.
- Isolate each character in the text in order to correctly identify it.
- Interpret the recognition results and resolve any ambiguities in order to reproduce the words accurately on the page.
- Format the results by examining and reproducing the font and page formatting information if desired.

AnyFont

AnyFont technology is the first true page-recognition approach to OCR. Rather than work from the matrix-matching approach, or from a machine-trained probability matrix analysis used by the dedicated OCR system providers, Caere takes an expert system approach to OCR that begins with the entire page.

The high-end systems that did omnifont OCR cost over $50,000 at the time their original algorithms were created. The main customers for such expensive systems were extremely text- intensive enterprises such as government agencies, insurance companies, and large law offices.

These customers shared a document pool that did not vary much in terms of fonts, point size, or page layout. The algorithms designed for these simple documents were insufficient for the needs of the general office where trade magazines, books, and desktop-published reports can have a wide variety of fonts and layout. Robert Noyce, the founder of Caere,

challenged his development team to create an OCR system that could read anything that may be in a typical office: even a magazine page with its mixture of fonts, text, and graphics.

Page Analysis

An 8 1/2" by 11" document at 300 dots per inch creates a 1.2 MB file. AnyFont first puts that entire file into RAM and looks at the complete page. From there, the algorithms look for and separate areas which may contain graphics by using image-processing techniques such as texture and density analysis and edge detection.

Next, the algorithm looks for columns of text and determines their size and shape. White lines between the text are discovered, with no preconceived rules regarding leading (the space between lines). This step allows OmniPage to recognize drop caps as the font size of each letter is not pre-determined. Finally, the spaces between the letters are found, and within this variable-sized zone lies a character of somewhere between 6 and 72 points. This zoom-in from the entire page technique allows OmniPage to recognize the widest variety of font sizes in the industry.

Character Experts

Each character box is sent to a team of 100 expert systems with each subsystem responsible for the identification of a single character. This is in marked contrast to other OCR technologies which are based upon a probability analysis of dots within a matrix field.

By the probability approach, the program is never certain if the character is an “a” or an “a,” but only the probability of each being the case. If the probability of it being an “a” is 89% and of being a “c” is 69%, then the “a” is reported. However, had the probability been 69% for a “c” and only 58% for an “a,” then the “c” would have been reported. In either case, it is a guess and suffers an inherent weakness of substitution errors. Products based on this technology rely heavily on dictionaries and language-usage probabilities to help determine which of their guesses belongs in a known word. This doesn’t help much with proper nouns, new terminology, or abbreviations.

With AnyFont’s 100 experts, each expert is responsible for identifying a character image. The first expert evaluates the image and decides if it is the character for which it is responsible. If it is not certain, it passes the character to the next expert. This continues until each character is recognized. No probabilities or guesses are used. This individual expert technology is theoretically the most accurate method possible of doing character recognition. It does not depend upon the averaging of a data base and the variations or guess work of probability pools. Each expert for

each character can be infinitely tuned and re-tuned as new fonts or new problems come up.

If there is a problem with “c”s and “e”s, additional tuning of those two experts is done until that one problem is resolved. To recognize a foreign language that has an “ä” as well as an “a,” another expert to identify the new character is added. This expert approach to recognition is what allows OmniPage to recognize more languages (13) than any other OCR package. It is this expert approach that provides OmniPage with so remarkably few substitution errors.

The inherent accuracy of the algorithm has always been the most important parameter at Caere. Experts provide that; however, they have two down sides. One is that they are remarkably difficult and time-intensive to program. Such an approach would be incredibly accurate for Kanji, but would take several hundred man-years to program the 5000 character experts needed! Machine-learned database probability pools or neural nets are the most practical approaches for such a language.

Self-Learning OCR

The other downside of experts is that they are very computer intensive, and therefore somewhat slow. One of the patents pending for Caere has to do with an accelerating, self-learning routine which allows each unique character to only have to be recognized once. From then on, the system will identify it as another “a” or “b” without having to reanalyze it each time with the experts. This accelerator technique makes OmniPage actually speed up as it reads a document. This technology, operating in true 32-bit mode, makes AnyFont the fastest omnifont OCR algorithms in the world, with speeds of up to 4000 words per minute attainable on the faster PCs.

Sometimes none of the experts are able to identify the character. This can happen with broken or overlapping characters. This is solved by AnyFont’s second pass. It can be seen on the screen as the light blue areas of the document image are painted a darker blue. The characters, or pieces of characters, that get past the experts are put in a separate buffer to be dealt with later. A series of very sophisticated routines come into play for splitting, combining, fragment analysis, fatting, thinning, and context checking. The quality and sophistication of these second pass routines provide greater recognition accuracy, even for very difficult problem characters and parts of characters. A third pass allows the Language Analyst to further refine accuracy.

AnyPage

AnyPage technology is Caere's proprietary dynamic auto-thresholding technology. It is in many ways similar to the HP AccuPage 2™ technology available with Hewlett-Packard scanners. AnyPage works with grayscale images. This dynamic thresholding technology requires a great deal of interaction between the scanner and the OCR software to vary the definition of "background" on a dynamic, real-time basis.

On a page with multiple shades of background, from a colored side bar in a magazine or newspaper, to a coffee stain on a business letter, AnyPage automatically varies which shades of gray will be considered "background" and which will be considered "type." AnyPage does this not just once for the entire page, or large blocks at a time, but over thousands of regions per page, sensing the optimum level for each region. The remaining levels are dropped out, leaving the image very clean and sharp and transforming it into a bi-modal black and white image to be delivered to the AnyFont OCR engine.

This technology allows OmniPage to recognize documents and parts of documents that other OCR packages literally can't even see. Black print on gray stationary, sidebars in the Wall Street Journal, and stained, yellowed, and faded documents. All these can now be seen, cleaned up, and recognized.

An important benefit to Dynamic Automatic Thresholding is that not only fringe or difficult documents need contrast adjustment; a variety of regular documents do too. With AnyPage operating with a sheet feeder, the user is assured that each document in the pile will be properly recognized, with no need to stop and change the scanner settings as the documents in the stack vary.

Compound Neural System

Just as our brain can recognize letters on a page even when part of the letter is missing or misprinted, Caere has developed a compound neural system that has "learned" to recognize characters even when the characters have been damaged.

This neural system is used for higher accuracy on characters that are distorted through printing and faxing and for text printed in small point sizes.

Caere's neural system consists of several neural networks whose components are modeled after the nerve cells in a brain. Each network consists of rows and columns of software-simulated neurons. Each neuron (Caere actually employs "perceptrons" as the active neurons in OmniPage) weighs evidence provided by the image and other neurons to

compute the probability of a specific character identification. The compound neural system compiles all the available evidence before making a character determination.

AnyFax

The AnyFax™ portion of the neural system is specially designed to work with low resolution fax images. It uses an image enhancement process that disconnects joined characters and reduces the jagged character edges created during the faxing process.

Fax transmissions often suffer from "noise" during transmission, which results in missing lines. Since fax signals are represented by a series of CCITT codes, AnyFax technology was developed to determine the missing CCITT codes and reconstruct them.

The Language Analyst

The Language Analyst is a feature that is available in both OmniPage and OmniPage Professional. The Language Analyst uses information about language context and usage rules to evaluate characters and words *during* the recognition process.

In addition to the character analysis techniques that OmniPage uses, the Language Analyst examines each entire word on the page during recognition and corrects any errors it finds. The Language Analyst refers to dictionaries with over 100,000 words in up to 11 languages to correct any errors in the recognition process as they occur.

When developing the Language Analyst, Caere had the benefit of the experience of its users from 35 countries using 16 different languages to provide examples of the most common OCR errors that needed correction.

Trigram Analysis

Trigram analysis looks at letter sequences in sets of three and determines the likelihood that those three letters would appear together in the language that was used for text recognition. For example, in English text, OmniPage will generally expect the letter u to follow the letter q and knows that qui is a more likely letter combination than qul.

3D OCR

3D OCR is a revolutionary new technology that is available to OmniPage Professional users. *3D* OCR is designed to provide precision accuracy for the most difficult-to-read documents such as multi-generational photocopies and pages with very small type. It does this by recognizing gray parts of images as well as bi-level, black and white parts.

Examination of the character's gray data, in addition to the black character image, provides a complete picture of the character's shades, shapes, and contours for the highest accuracy possible.

For example, imagine typing on a typewriter with a ribbon that was running out of ink. Some of the characters on the page will be illegible; for instance, the letter n might look like the letter r. The missing section might appear very lightly inked; *3D* OCR evaluates the grayscale information of that lightly inked area and correctly recognizes the character as an n.

Or, imagine applying too much ink to a rubber stamp. When you apply the stamp to a piece of paper, the words may be completely blurred. *3D* OCR evaluates the grayscale information of that extra inked area and decides that compared to the darkness of most of the character, the blurred area is probably a smudge. It the evaluates only the relevant parts of the character for recognition.

True Page™

OmniPage Professional allows you to maintain the page layout of your documents – with both text and images – throughout the recognition process in their original full-page formats. This feature is called True Page recognition.

When you select the True Page option, all the page attributes are displayed, complete with columns, typestyles, font characteristics, and graphics as they appeared on the original page. See Chapter 7, How to Use True Page, for detailed information.