

OmniPage Limited Edition

Installation and Setup

This section describes how to install OmniPage Limited Edition (referred to hereafter as OmniPage Ltd. Ed.) and begin using it. It contains the following topics:

- What is Optical Character Recognition (OCR)?
- System Requirements
- Scanner Requirements
- Installing OmniPage Ltd. Ed.

What is Optical Character Recognition (OCR)?

Optical character recognition (OCR) is the process of converting an image to computer-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 a TIFF or PICT file (for example, a received fax file can be saved to one of those formats and then loaded into OmniPage Ltd. Ed. for OCR).

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 Ltd. Ed. looks for and defines characters on the image to produce editable text. You can then export the recognized text

from OmniPage Ltd. Ed. to a variety of word-processing, page layout, and spreadsheet programs.

OCR is also referred to as *text* or *page recognition*, or just *recognition*.

System Requirements

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

- Macintosh (68020 or greater), Power Macintosh, or Macintosh-compatible computer
- System 7.0 or later
- Free memory of at least 5MB RAM
- Hard disk space of at least 7MB



If you use System 7.0 or 7.1, remove the Dragging Enabler file from your Extensions folder. OmniPage Ltd. Ed. installs an AppleGuide enabler from Apple that replaces the Dragging Enabler file.

Scanner Requirements

The scanner driver supplied by the scanner manufacturer, if any, must be installed on your system according to the manufacturer's instructions.



Make sure your scanner is set up and connected properly according to the manufacturer's instructions.

OmniPage Ltd. Ed. can open line art and grayscale TIFF and PICT image files produced by other scanners. TIFF files must be black-and-white or grayscale, and 200, 300, or 400 dpi. A grayscale image will be imported as a black-and-white image.

Installing OmniPage Ltd. Ed.



Before installing OmniPage Ltd. Ed., restart your Macintosh while holding down the Shift key to turn off all extensions, including virus-checking software. Extensions may interfere with installation. Or, you can use the Extensions Manager control panel to turn off extensions.

- 1 With all extensions off, insert OmniPage Limited Edition disk #1 in your computer's floppy drive, or place the CD-ROM in your CD-ROM drive.
- 2 Double-click to open the disk if it does not open automatically.
- 3 Double-click the Installer icon.
If you are installing from disks, follow installation instructions and insert the other program disks as prompted.

Save your registration code in writing as well. You will need it again if you have to reinstall OmniPage Ltd. Ed. for any reason.

Scanning and Recognizing Pages

This section describes how to use OmniPage Ltd. Ed. to convert a scanned page or an image file to editable text that you can use in other applications. It also describes the various settings and functions OmniPage Ltd. Ed. provides.

This section contains the following:

- Selecting a Scanner
- Scanning a Page
- Loading an Image File
- Using Direct Input
- Manual Zoning
- The Image Window
- The AutoOCR™ Toolbar
- The Settings Panel

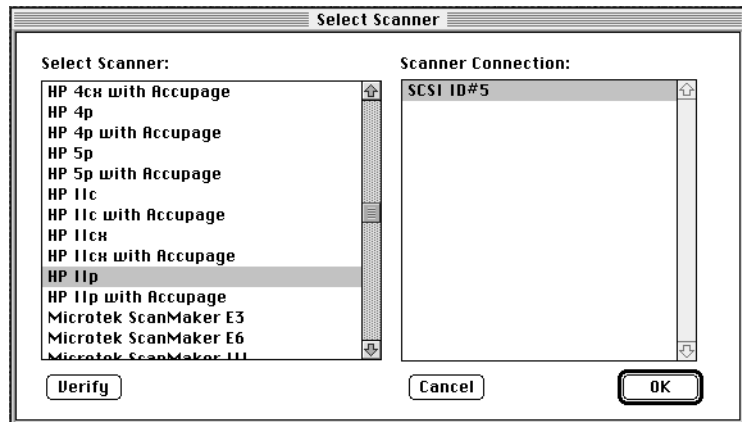


If OmniPage Direct is installed on your system, remove it before using the Direct Input feature in OmniPage Ltd. Ed. OmniPage Direct may interfere with using Direct Input. See “Using Direct Input” on page 9 for more information on this feature.

Selecting a Scanner

This section describes how to select or change a scanner so that it will communicate with OmniPage Ltd. Ed.

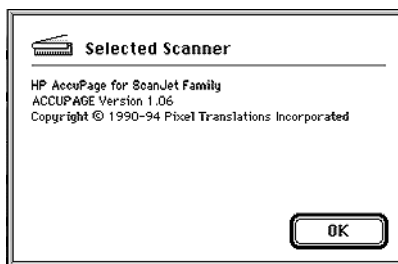
- 1 Choose *Select Scanner* in the Settings menu.
The Select Scanner window appears.
- 2 From the left panel of the Select Scanner window, highlight the description of your particular scanner.
The Select Scanner window shown is an example. Additional information or choices related to your scanner may appear in the right pane.



If you click the Verify button, OmniPage Ltd. Ed. will attempt to communicate with your scanner.

If communication is successful, the Verify Scanner window appears, which describes the scanner or scanner family you selected and the software-driver extension you have installed.

Unsuccessful communication results in an error message with corrective suggestions.



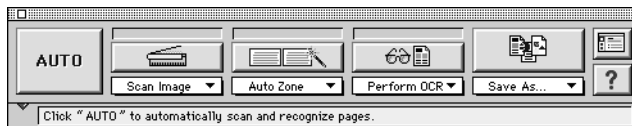
- 3 Click OK on the Verify Scanner window and the Select Scanner window.
Both windows close.

Scanning a Page

This section describes how to scan, recognize, and save a page automatically.



- 1 Double-click the OmniPage Limited Edition icon in the OmniPage Limited Edition folder on your hard drive.
The AutoOCR™ toolbar appears with the default settings selected the first time you launch the program.



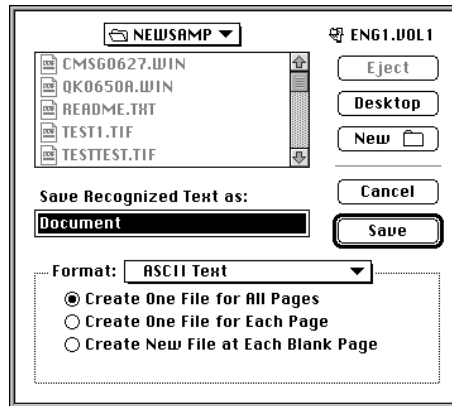
- 2 Select these settings under each button's pop-up menu if they are not selected already: *Scan Image*, *Auto Zone*, and *Save As....* (The third button's pop-up menu only contains *Perform OCR*.)
- 3 Place a page in your scanner, making sure that it is aligned correctly.
Use a single-column document such as a memo for best results.
- 4 Click the *AUTO* button in the toolbar to begin automatic processing.
 - Your scanner scans the page and the image appears in the Image window.
 - OmniPage Ltd. Ed. draws a single zone around the entire page.



- OmniPage Ltd. Ed. makes three OCR passes over the image as it analyzes and recognizes the text.
- The recognized text appears in the Text window.



- 5 Click the Save As... button in the toolbar to open the Save As dialog box.



- 6 Type a name for your document in the *Save Recognized Text* as box.
- 7 Select a format for your document in the *Format* pop-up menu.
- 8 Click *Save*.
- 9 You can now open and edit your document in the appropriate program.

If you selected *Word 6.0* in the *Format* pop-up menu, for example, you can open the saved document in Microsoft Word or any other program that supports Word files. If you saved it as *Uncompressed TIFF* (or another type of TIFF), you can open it in any graphics program that supports TIFF files.

Loading an Image File

This section describes how to load and recognize supported TIFF and PICT files. You may have a received fax file, for example, that was saved in TIFF format. You can turn this fax image into editable text by recognizing it in OmniPage Ltd. Ed.



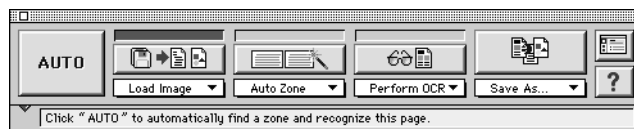
TIFF files must be black-and-white or grayscale, and 200, 300, or 400 dpi. A grayscale image will be imported as a black-and-white image.

To load and recognize an image file:

- 1 Launch OmniPage Ltd. Ed. as described in the previous section if it is not open already.

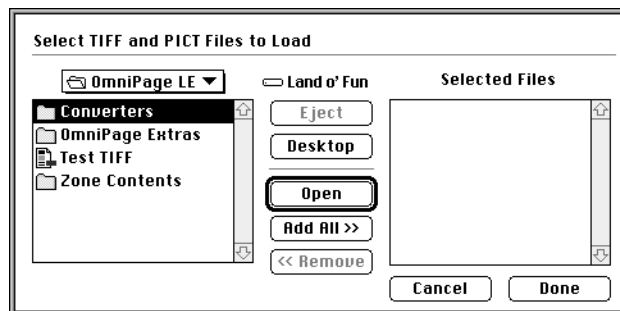
The toolbar appears.

- 2 Select these settings under each toolbar button's pop-up menu: *Load Image*, *Auto Zone*, and *Save As....* (The third button's pop-up menu only contains *Perform OCR*.).



- 3 Click the *AUTO* button in the toolbar to begin automatic processing.

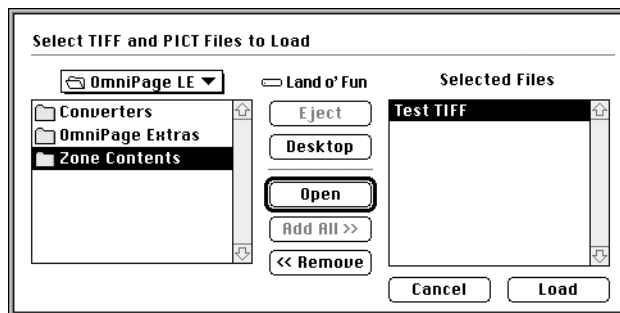
The Load Image dialog box appears.



- 4 Locate and select the *Test TIFF* file in your OmniPage Limited Edition folder, or any other TIFF or PICT file you want to recognize.

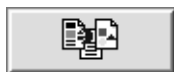
The *Open* button becomes the *Add>>* button.

- 5 Click *Add>>* to move the selected file into the *Selected Files* list box on the right.

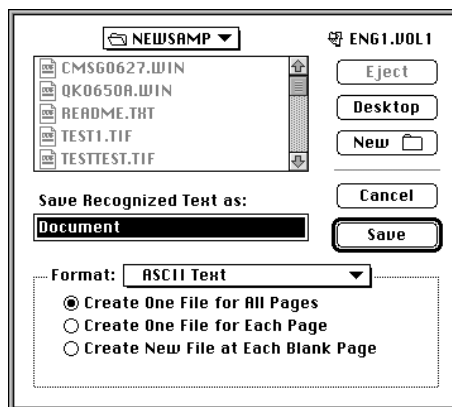


You can select up to 20 files.

- 6 Click *Load*.
 - The image appears in the Image window and OmniPage Ltd. Ed. draws a single zone around the entire page.
 - OmniPage Ltd. Ed. makes three OCR passes over the image as it analyzes and recognizes the text.
 - The recognized text appears in the Text window.



- 7 Click the *Save As...* button in the toolbar to open the *Save As* dialog box.



- 8 Follow steps 6–8 on page 6 to save the recognized text.

Using Direct Input

This section explains how to use Direct Input to initiate OCR processing from an open application and paste recognized text directly into that application. First, please read the warning on page 2-1 regarding OmniPage Direct.

To use Direct Input:

- 1 Align your document(s) in your scanner or automatic document feeder (ADF) if you plan to scan.
- 2 Open or switch to any application that supports pasting from the Clipboard, such as a word-processing program.
If you do not have any application open, OmniPage Ltd. Ed. will automatically paste recognized text to the Clipboard.
- 3 Choose *OmniPage Ltd Direct Input* in the Apple menu.
 - OmniPage Ltd. Ed. launches in Direct Input mode if it was not already open.
 - OmniPage Ltd. Ed. switches to Direct Input mode if it was already open.



The toolbar appears with *Auto Paste* as the default selection in the Save As button's pop-up menu.

- 4 Select the appropriate settings for your document under the other buttons.
- 5 Click the *AUTO* button to begin processing.
OmniPage Ltd. Ed. processes the page according to the selected settings and pastes the recognized text into your other open application.



Direct Input works best with a small number of scanned pages or images. If you have more than five pages, launch OmniPage Ltd. Ed. and use its *Save As...* command to export recognized text instead. Your Macintosh must have enough memory to run both OmniPage Ltd. Ed. and the other application.

After processing, Direct Input mode closes and OmniPage Ltd. Ed. quits if it was launched via Direct Input.

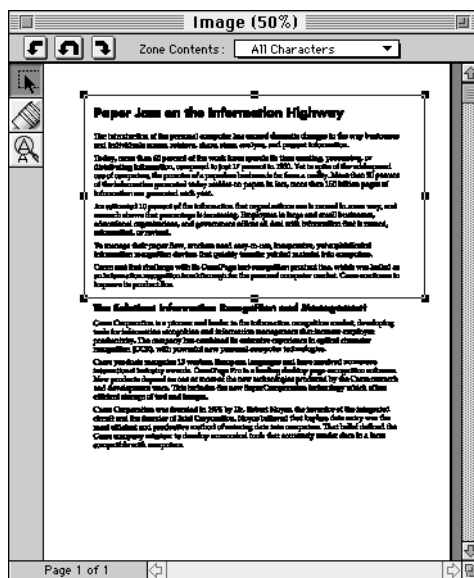
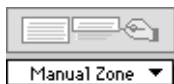
OmniPage Ltd. Ed. appears again if it was running when Direct Input was selected.

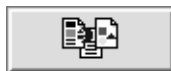
Manual Zoning

You can select which portion of a page to recognize by using *manual*, rather than *automatic* zoning. This is useful if you need just part of a page. Recognizing part of a page takes less time than recognizing the whole page.

To use manual zoning:

- 1 Launch OmniPage Ltd. Ed. if it is not open already.
- 2 Choose *Manual Zone* in the pop-up menu under the Zone button in the toolbar.
- 3 Select the appropriate settings for your document under the other buttons.
- 4 Click the *AUTO* button in the toolbar to scan a document or load an image file, depending on what you selected under the Image button.
See “Scanning a Page” on page 5 or “Loading an Image File” on page 6 for more information.
The image appears in the Image window and your cursor turns into a crosshair over the page.
- 5 Hold down the mouse button and draw a zone around the text to recognize.



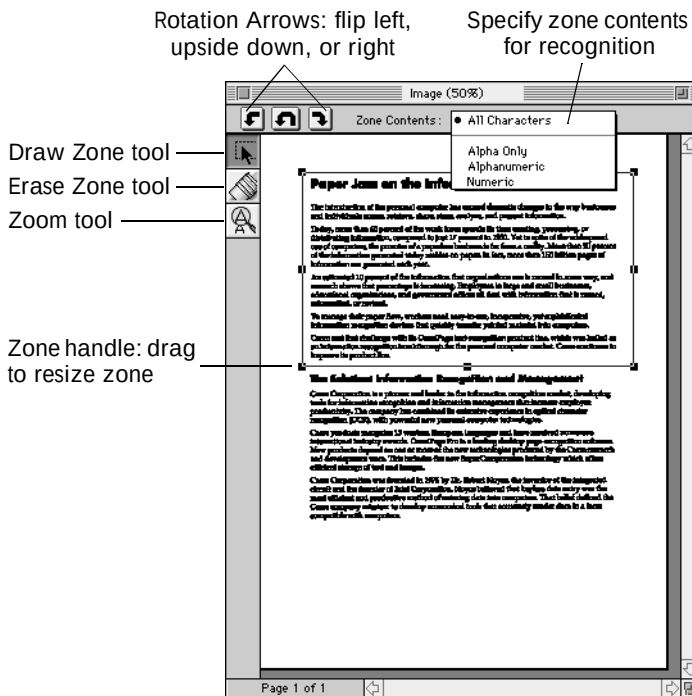


- 6 Click the *AUTO* button in the toolbar to finish OCR processing.
 - OmniPage Ltd. Ed. makes three OCR passes over the image as it analyzes and recognizes the text.
 - The recognized text appears in the Text window.
- 7 Click the Save As... button in the toolbar to open the Save As dialog box.
- 8 Follow steps 5–8 on page 6 to save the recognized text.

See “The Image Window” on page 11 for more information on this window.

The Image Window

This section explains the tools and options in the Image window.



Draw Zone tool: Draw a zone around the portion of the page to recognize. Drag a zone handle to resize the zone.

Erase Zone tool: Click the zone to erase it.

Zoom tool: Click the image to zoom in. Option-click to zoom out.

Rotation arrows: Use to flip the image left, upside down, or right.

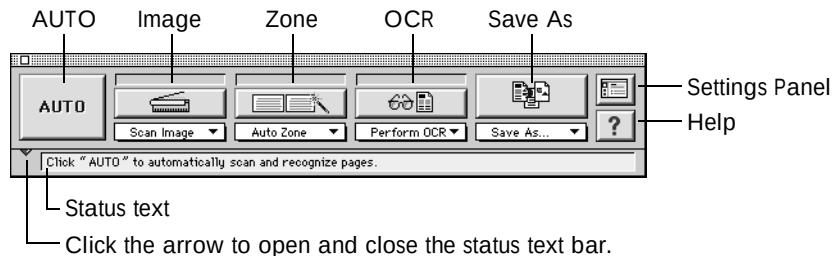
Zone Contents menu: Select *All Characters* to attempt to recognize any printed character in the selected language. (Choose *Select Languages...* in the Settings menu to select one or more languages for a document.)

- Select *Alpha Only* to attempt to recognize just letters of the alphabet for the selected language.
- Select *Alphanumeric* to attempt to recognize numbers and letters of the alphabet for the selected language.
- Select *Numeric* to attempt to recognize numbers only.

The AutoOCR™ Toolbar

The AutoOCR toolbar appears whenever you launch OmniPage Ltd. Ed. It contains seven buttons. The five larger buttons are the *process* buttons.

You can click each button individually, or you can click *AUTO* to automate processing as described in “Scanning a Page” on page 5 and “Loading an Image File” on page 6.



- **AUTO button:** Click this button to automate OCR. OmniPage Ltd. Ed. performs each processing stage selected (except *Save As...*) without you having to click the individual process buttons. You can also click *AUTO* to perform the next processing stage, such as recognizing text after drawing a manual zone.
- **Image button:** Select *Scan Page* in the pop-up menu to scan a hard-copy page. Select *Load Image* to load a PICT or TIFF file.
- **Zone button:** Select *Auto Zone* in the pop-up menu to have OmniPage Ltd. Ed. automatically zone the page before recognition. Select *Manual Zone* to draw your own zone anywhere on the image.
- **OCR button:** Click this button to begin OCR.

- **Save As button:** Select *Save As...* in the pop-up menu to save your recognized text to disk. Select *To Clipboard* to paste the text to the Clipboard. Select *Auto Paste* to paste the text automatically into another program's open document without having to switch to that document.

The Auto Paste command is only enabled in Direct Input mode. See "Using Direct Input" on page 9 for information.

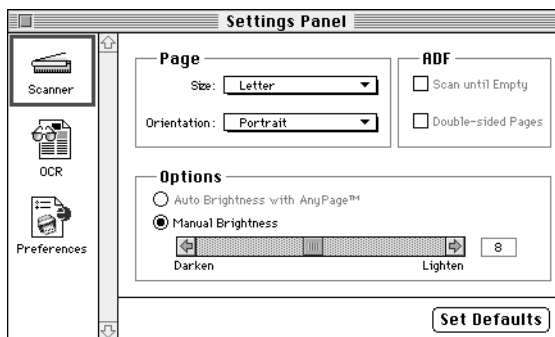
- **Settings Panel button:** Click this button to open the Settings Panel and select Scanner, OCR, and Preferences options. See "The Settings Panel" on page 13 for more information.
- **Help button:** Click this button to open the online help system.

The Settings Panel

Use the Settings Panel to select Scanner, OCR, and Preferences options.



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



Scanner Settings

Size

- Select *Letter* for 8.5" by 11" size pages.
- Select *A4* for 21 cm by 29.7 cm European-size pages.
- Select *Legal* for 8.5" by 14" size pages.

Orientation

- Select *Portrait* for a vertically-oriented page.
- Select *Landscape* for a horizontally-oriented page.

- Select *Flipped* to automatically rotate a portrait page image 180 degrees.
- Select *Flipscape* to automatically rotate a landscape page image 180 degrees.

Flipped and *Flipscape* options are useful if you scan pages in a book and have trouble positioning the book correctly.

ADF

These options are only available if your scanner has an automatic document feeder (ADF).

- Select *Scan until Empty* to scan multiple pages in the ADF when you click the *AUTO* button.
- Select *Double-sided Pages* to scan pages printed on both sides when you click the *AUTO* button.

OmniPage Pro processes the entire batch, and then prompts you to turn it over to process the reverse sides.

Options

- Select *Auto Brightness* for automatic-brightness correction of each scan. This setting is especially useful for: poor quality-pages; pages with very small type; or, pages with text on colored or shaded backgrounds. The default option is to use Caere's AnyPage™ brightness correction. Some scanners support other options (once properly installed and selected) and those options will be available via a pop-up menu here.
- Select *Manual Brightness* for crisp dark text on a white background. Use the slider to adjust brightness if the quality of your recognized documents is poor.

OCR Settings

Input Options

- Select *Normal* if the page has conventionally printed text characters (such as this manual).
- 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 on the page 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 both OCR-A and a conventional font, select *Normal* for faster recognition.

Output Text Options

- Select *Use Hard Carriage Returns* to place a return after each line of recognized text instead of wrapping each line.
- Select *Replace Each Tab with a Space* to replace areas between text of approximately five or more spaces 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.

Output Font and Size

In the *Font* and *Size* pop-up menus, select a font and point size for the recognized text.

Preferences Settings

Image Input Options

Select an orientation for the images you load. See “Orientation” on page 13 for detailed information on these options.

Preferences

Unrecognizable characters are represented by a reject character in the Text window. The default character is a tilde (~). You can type any character you want in the *Reject Character* box.

OmniPage Pro — Why Upgrade

OmniPage Pro is the most powerful OCR software available for the Power Macintosh. Use it to recognize more document types than any other OCR software: faxes, stacks of double-sided pages, spreadsheets, magazine or newspaper pages, legal documents, and so on.

OmniPage Pro Features

OmniPage Pro contains the following advanced features.

- **Accuracy:** OmniPage Pro makes a quantum leap in OCR accuracy. You can now perform OCR on more types of documents faster and with fewer errors. It is **over 99% accurate** on laser-quality documents that have standard fonts. OmniPage Pro also recognizes difficult documents like faxes and photocopies.
- **The Language Analyst®:** A Caere technology that uses supplied dictionary and linguistic data to apply context and usage rules. Caere's Language Analyst evaluates text and automatically corrects likely errors during recognition.
- **Color Scanning:** Color images can be scanned (using supported scanners) and saved. Graphics are retained in word processor formats.
- **True Page®:** Allows you to retain as much or as little of page layout, graphics, and formatting in a recognized document.
- **Multiple Zones:** Recognize complex layouts and tables. You can create up to 64 zones, specify the areas to recognize, or specify how to recognize zone contents (as text, numbers, or graphics), and in what order to recognize them.
- **Interactive Check Recognition:** Verify spelling and recognition with OmniPage Pro's spell-checker.
- **Verification Window:** Compare text in the recognized document with the original page image for quick and easy proofreading.
- **Train OCR:** Teach the program to recognize unusual or hard-to-read characters.
- **Deferred Processing:** Save time by scheduling OCR at your convenience.
- **Zone Templates:** Create, save, and load templates for faster, more accurate recognition of standardized documents.

OmniPage Pro 8 for the Macintosh will only operate on Power PCs.

It will not operate on 68K Macintoshes.

- **Upgrading to OmniPage Pro**

- 1 Choose *Upgrade OmniPage Limited Edition...* in the Apple menu to open the Upgrade dialog box.
- 2 Select your country in the pop-up menu.
A phone number appears to the right of the country menu.
- 3 Call this number to order OmniPage Pro for a special low price.

Technical Information

This section describes some basic troubleshooting techniques and how to improve performance.

This section contains the following sections:

- Troubleshooting
- Tips for Improved Performance

Troubleshooting

This section gives an overview of possible solutions to apply if you experience problems with OmniPage Ltd. Ed.

Try These First

- Turn off your computer, restart your scanner, and then restart your computer.
- Restart your computer with extensions off and then try scanning again.
- Reinstall OmniPage Ltd. Ed.

System Setup Checklist

- Make sure that your system meets all requirements as listed in “System Requirements” on page 1-6.
- Make sure that your scanner is plugged in and turned on, and that all cable connections are secure.
- Fix Macintosh problems before using OmniPage Ltd. Ed. again.
- Run virus-checking software regularly.

- Scan your disk for problems occasionally: use a utility that checks and diagnoses damaged disks and files, such as MacTools™ or Norton Utilities™.
- Defragment your hard disk occasionally with a program such as Norton Utilities or MacTools.

System Hang During Scan

A system hang during a scan could have several possible causes:

- A SCSI termination problem.
Make sure the scanner is terminated properly. See your scanner documentation. If you have several peripheral devices attached in a chain to the SCSI terminal, all devices must be turned on and the last device must also be terminated properly (unless it is terminated internally).
- A memory-related problem.
Try closing open windows and applications to free memory.

Low Memory

Low memory can cause error messages, slow performance or system hangs.

Choose *About This Macintosh...* in the Apple menu to check the amount of free memory. You may need to reconfigure your Macintosh for additional RAM.

Virtual Memory

Check your Macintosh documentation for information on using virtual memory. Virtual memory is hard disk space used temporarily as free RAM by your computer. Performance is slower with virtual memory.

Increasing Available Memory

Try closing open applications and windows to free up memory. Memory can become *fragmented* if many applications have been open at different times. Try restarting your computer to defragment the memory.

The best solution for low memory is to buy and install more RAM.

Low Disk Space

Make sure there is sufficient disk space for the temporary files that OmniPage Ltd. Ed. creates during scanning and OCR. Your disk should have at least 7MB of free disk space for the best performance (12MB is recommended). Otherwise, files may be fragmented (that is, written in small blocks in many different places on your disk), which slows the OCR process.

Free disk space by removing unneeded files or by copying less frequently used files to a floppy disk, an external hard drive, or another storage device.

Tips for Improved Performance

This section gives tips on achieving the best recognition accuracy possible.

Document Quality

OmniPage Ltd. Ed. recognizes characters in almost any font from 6 to 72 points in size. However, keep the following in mind:

- The print should be reasonably clean and crisp. Black text on a white background is best. Characters must be distinct: separated from one another and not blurry 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 Ltd. Ed. may not recognize the *Zapf Chancery* font accurately.
- Documents with underlined text should have the lines placed below the text and not touching it for best results. It is difficult to recognize underlined text because the underline overprints the descenders on the letters *g*, *j*, *p*, *q*, and *y*, changing their shape.
- Ask senders to fax documents in “Fine” or “Best” mode for the highest resolution.

Scanning Angle

Make sure that each document is positioned correctly in your scanner and is not crooked. Adjust the page and rescan it if you see numerous recognition errors due to skewing.

Scanner Glass Clarity

The sheet of glass on the flatbed of the scanner must be clean. If it gets dirty, wipe it gently with a soft, damp, lint-free cloth or tissue. Be sure that it is completely dry before you put pages on it.

See your scanner documentation for more information on proper scanner maintenance.

Paper Transparency

Some paper is thin enough that the scanner sees text printed on the other side of a two-sided page. To correct this problem, place a black piece of paper between the page and the lid of the scanner.

Language Selection

If a foreign-language document produces numerous OCR errors, make sure you selected the correct language for the document.

Choose *Select Languages...* in the Settings menu. Select the correct language in the Select Languages dialog box and click *OK*. Then re-recognize the document.

Using HP AccuPage™

HP AccuPage is a technology developed and licensed by Hewlett-Packard. If your scanner supports HP AccuPage, you can use it to:

- Improve the recognition of pages printed on colored or shaded backgrounds, or on smudged paper.
- Improve the recognition of pages with very small type.
- Improve the recognition of poor-quality pages such as second-generation copies or low-resolution faxes.

To use HP AccuPage when scanning:



- 1 Click the Settings Panel button in the toolbar to open the Settings Panel.
- 2 Click the Scanner icon.
- 3 Select *Auto Brightness with HP AccuPage*.



Try selecting *Manual Brightness* and adjusting the darkness or lightness of scanning if HP AccuPage does not produce the expected quality.

Zone Contents

You can specify the zone *content* of a manually drawn zone for better and faster recognition.

For example, if you scanned a spreadsheet that consisted entirely of numbers, you could choose *Numeric* in the Zone Contents drop-down list before recognition. OmniPage Ltd. Ed. would only look for and recognize numbers during recognition instead of possibly mistaking the number 50 for the word S0.