



Introduction to OmniPage Pro OCR Training at the Institute

Presented by
Information Technology Group
Institute for Advanced Study
Einstein Drive
Princeton, NJ 08540
(609) 734-8044 ♦ helpdesk@ias.edu

12/1/2008

Table of Contents

Introduction	1
Getting Help	2
Introduction to OCR using OmniPage Pro	3
What is Optical Character Recognition (OCR)?	3
OmniPage Pro's OCR capabilities	3
Working with documents in OmniPage Pro	4
The OmniPage Pro Environment	5
Launching OmniPage Pro	5
The OmniPage Pro Desktop	5
Basic OmniPage Pro OCR Processing Steps	11
Processing Documents	12
Processing methods	12
Processing automatically	13
Processing manually	14
Defining the source of page images	15
User dictionaries	18
Text and image editing	19
Saving and exporting	20
Quick Start Guide	23
Loading and recognizing sample image files	23
Scanning and recognizing a single page	24

Introduction

OmniPage Pro is an Optical Character Recognition (OCR) scanning program designed to transform text and/or images from scanned documents into editable text for use in other software applications such as Microsoft Word. The benefits of this application's latest release are many:

- Greater character recognition accuracy.
- Improved page layout recognition.
- Portable Document File (PDF) file capability allows the user to easily create PDF files from any scanned document.
- Language support for over 100 languages.
- Built-in text editor.

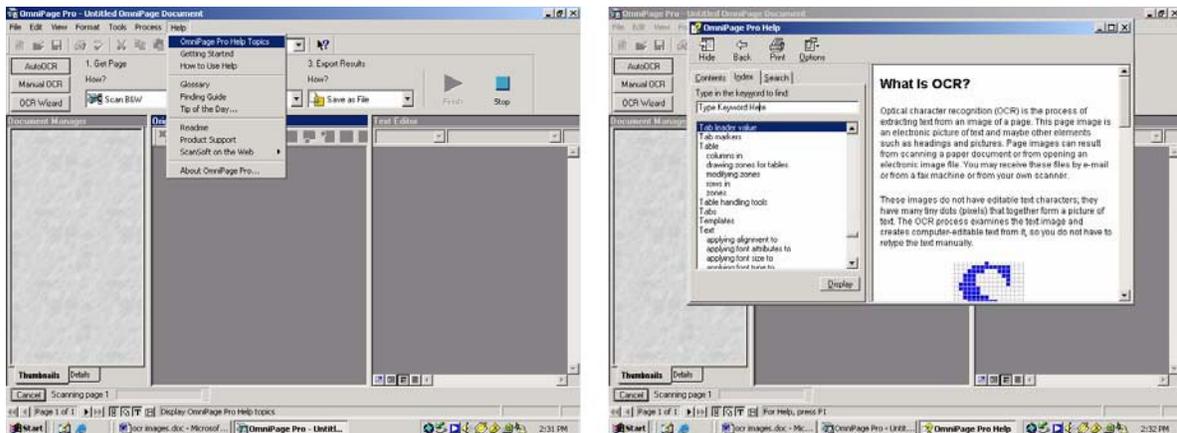
This training manual was created based on information appearing in the *OmniPage Pro User's Guide*. This training manual will assist you with the following:

- Learning more about Optical Character Recognition (OCR).
- Becoming familiar with the features of OmniPage Pro.
- Performing OCR processing on image files and scanned documents.
- Defining page layouts and zones.
- Saving your scanned results to different file formats.

Getting Help

If you experience problems using OmniPage Pro, you can obtain help for the following two sources:

OmniPage Pro's Online Help The online Help contains information on features, settings and procedures. To access this feature, choose **OmniPage Pro Help Topics** from the **Help** menu bar option. A table of contents will be displayed to assist you with locating the information you are seeking. Pressing **Shift + F1** will also display the online Help dialog window.

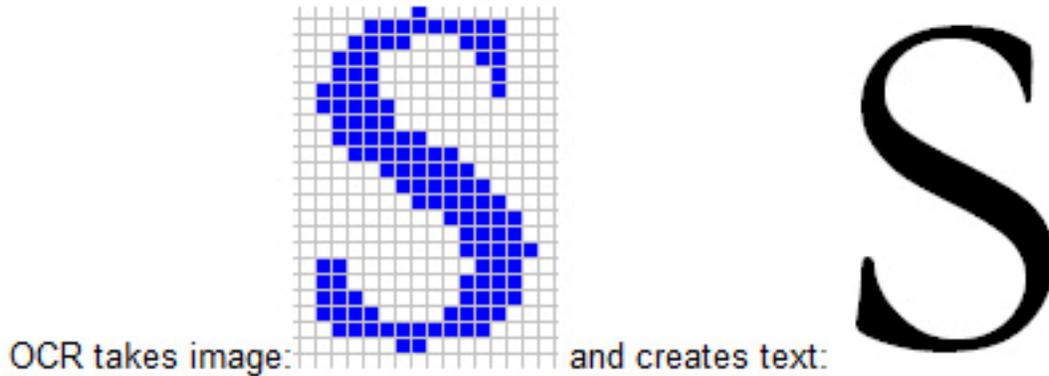


Information Technology Group Help Desk The IAS Information Technology Group Help Desk can be reached Monday through Friday between the hours of 8:00 a.m. and 5:00 p.m. by telephone (extension 8044) or e-mail (helpdesk@ias.edu)

Introduction to OCR using OmniPage Pro

What is Optical Character Recognition (OCR)?

Optical character recognition (OCR) is the process of extracting text from an image file. This image can result from scanning a paper document or opening an electronic image file. Images do not have editable text characters; they have many tiny dots (pixels) that together form character shapes. These present a picture of the text on a page. During OCR, OmniPage Pro analyzes the character shapes in an image and defines solutions to produce editable text. After OCR, you can save the resulting text to a variety of word-processing, desktop publishing or spreadsheet applications.



OmniPage Pro's OCR capabilities

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

Element	Description
Graphics	Photos, logos, and drawings are examples of graphics.
Form elements	Checkboxes, radio buttons, text fields.
Text Formatting	Font types, sizes and styles (such as bold, italic and underlines) are examples of character formatting. Indents, tabs, margins and line spacing are examples of paragraph formatting.
Page formatting	Column structure, table formats, and placement of graphics and headings are examples of page formatting.

Note: OmniPage Pro only recognizes machine-generated characters such as offset or laser-printed or typewritten text. However, it can retain handwritten text, such as a signature, as a graphic. For optimum scanning results, please be sure to use high quality printed materials featuring crisp text. Low quality printed materials (i.e., faded text, photocopies, sheets folded in half, etc.) will not produce reliable OCR results.

Working with documents in OmniPage Pro

A document in OmniPage Pro consists of one image for each document page. OmniPage Pro handles documents one page at a time. When you acquire your first image (from a scanner or from an image file) a new document is started. Further acquired images can be added to the same document. After you perform OCR, the document will also contain recognized text, displayed in the **Text Editor**, possibly along with graphics, form elements and tables. Once this is done, you will be able to save and close the document.

The OmniPage Pro Environment

Launching OmniPage Pro

To start OmniPage Pro, perform the following:

1. Click on the **Start** button.
2. Select All **Programs**.
3. Select **ScanSoft OmniPage 16**.
4. Click on the **OmniPage Pro 16** icon.

Or

Double-click on the **OmniPage Pro** icon on the desktop.



OmniPage Pro.Ink

Once the application is opened, the OmniPage Pro title screen and desktop will be displayed.

The OmniPage Pro Desktop

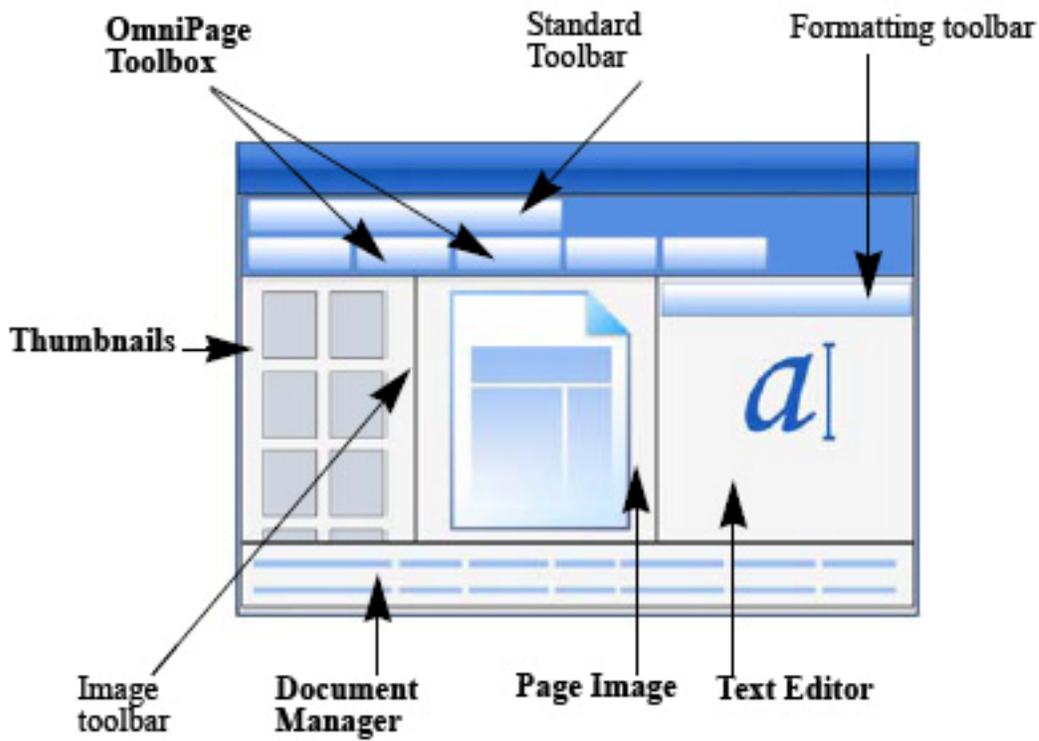
OmniPage comes with three different views to suit your task the best.

- Classic View - This view has a similar look and feel to previous versions of OmniPage.
- Flexible View - This view is a new alternate layout of the OmniPage function panels stacked in a tabbed view to give each panel more space.
- QuickConvert View - This view is designed for quick and easy document conversion without having to learn a lot. The most important conversion options are clearly visible on one screen.

Use the Windows menu to switch between views and to save your own custom view. For a custom view, arrange the panels and toolbars as you wish, then choose Window > Custom Views > Manage. Click Add and name your view. Your screen layouts will be displayed in the Custom Views submenu with a checkmark beside the active one.

Classic View

In Classic View, the OmniPage Desktop has four main working areas, separated by splitters: the Document Manager, the Page Image, Thumbnails and the Text Editor. The Page Image has an Image toolbar and the Text Editor has a Formatting toolbar.



OmniPage toolbox: This Toolbox lets you drive the processing.



Buttons

<p>Start button</p> 	<p>Initiates automatic processing: acquiring images, performing OCR (auto-zoning, recognition and, if requested, proofing) and exporting according to the current settings. Starts a selected workflow. The button has a different appearance depending on the selection in the Workflow drop-down list.</p>
<p>Get Page button</p> 	<p>Scans pages (in color, in gray or BW) or loads images. The button has different pictures according to the document source selected in the Get Page drop-down list.</p>
<p>Perform OCR button</p> 	<p>Initiates text recognition. For images without zones also initiates auto-zoning prior to text recognition. The button has different pictures according to the setting selected in the Layout Description drop-down list .</p>
<p>Export Results button</p> 	<p>Serves for saving the recognition results in a specified file format, for sending as mail or to other targets, or for copying to the Clipboard. The button has different pictures according to the destination selected in the Export Result drop-down list.</p>

Thumbnails panel: This displays page thumbnails.

Document Manager: This provides an overview of your document with a table. Each row represents one page. Columns present statistical or status information for each page, and (where appropriate) document totals.

Page Image: This displays the image of the current page, together with its zones. When a page is displayed, the Image toolbar is available.

Text Editor: This displays the recognition results from the current page.

Flexible View

Use this view to set up the OmniPage workspace so that it fits your task optimally. Suggested scenarios:

Maximizing workspace (single screen)



Load a document. Open the panels you want to use. Grab them by their captions one by one, and drag them so that they dock behind the active one as tabs. You can also dock online Help to avoid handling two separate windows

Working with recognition results (single screen)



Load a document and have it recognized. Close all panels except the Document Manager and the Text Editor. Maximize both horizontally, scale down the Document Manager and dock it to the top or bottom. You can now step through the pages double-clicking them one by one in the Document Manager, inspecting recognition results in the Text Editor. The number of suspect words and reject characters in the Document Manager will help you identify problematic pages.

Handling large documents (dual-screen)



Load the document you want to work on. Move its Thumbnail View to your second monitor and maximize it for a large scale overview of your document and far more space for thumbnail operations.

Verifying (dual-screen)

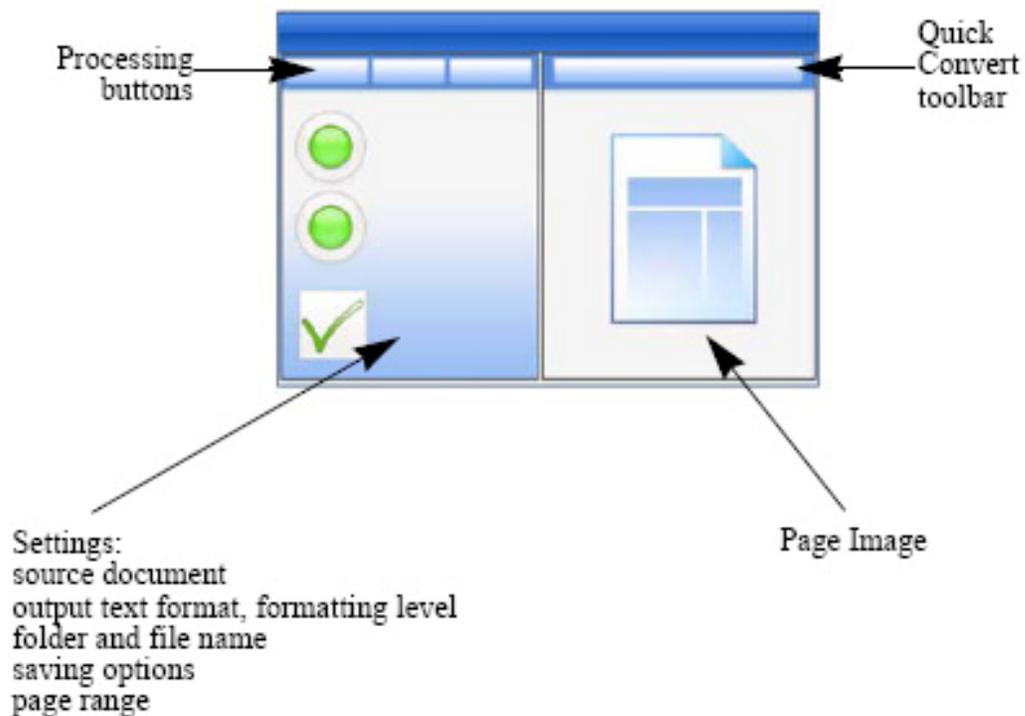


Place the Page Image on one screen and the Text Editor on the other. This gives you more space for editing and proofing. The Page Image is always available for verifying recognition and for performing on-the-fly zoning and editing.

The scenarios presented above are only examples to give you an idea of what you can do in Flexible View.

QuickConvert View

Use the QuickConvert View for fast recognition and saving. You can switch to Quick View only when you have no opened document and it can handle only one document at a time.



The Toolbars

The program has eleven main toolbars. Use the View menu to show, hide or customize them. Status bar texts at the bottom edge of the OmniPage program window explain the purpose of all tools.

Standard toolbar: Performs basic functions.



Image toolbar: Performs image, zoning and table operations. Three of its tool groups can now be handled separately (mini-toolbars):

- Zones toolbar: Offers zoning tools.
- Rotate toolbar: Provides rotating tools.
- Table toolbar: Inserts moves and removes row and column dividers.



Tools	Tool names
	Select zone tool.
	Draw zone tools: Draw text zone tool , Draw table zone tool and Draw graphic zone tool .
	Zoom in tool , Zoom out tool
	Rotate image tools: Rotate right , Rotate left , Rotate 180 degrees .
	Show / hide Overview tool
	First Image , Previous Image , Go to Page , Next Image , Last Image . These navigation tools are disabled if the document source is Load Files or Load Digital Camera Files.

Formatting toolbar: Formats recognized text in the Text Editor.

Verifier toolbar: Controls the location and appearance of the verifier.

Reorder toolbar: Modifies the order of elements in recognized pages.

Mark Text toolbar: Performs text marking and redacting.

Form Drawing toolbar: Creates new form elements.

Form Arrangement toolbar: Arranges and aligns form elements. All toolbars can be moved and customized in each view to your particular needs, including use of a secondary monitor.

Program Panels

OmniPage has six panels that can be handled (docked, floated, resized) separately: Thumbnails, Page Image, Text Editor, Document Manager, Workflow Status, and Online Help.

To float a panel anywhere on the screen, keep CTRL pushed while dragging. To dock it, drag the panel over the OmniPage main window, hold down the left mouse button and start pressing space to see all possible docking positions. To select a given position, release the mouse button.

Basic OmniPage Pro OCR Processing Steps

There are three ways of handling documents: with automatic, manual or workflow processing. The basic steps for all processing methods are broadly the same:



1. Bring a set of images into OmniPage. You can scan a paper document with or without an Automatic Document Feeder (ADF) or load one or more image files.



2. Perform OCR to generate editable text. After OCR, you can check and correct errors in the document using the OCR Proofreader and edit the document in the Text Editor.



3. Export the document to the desired location. You can save your document to a specified file name and type, place it on the Clipboard, send it as a mail attachment or publish it. You can save the same document repeatedly to different destinations, different file types, with different settings and levels of formatting.

Using OmniPage, you can choose from the following processing methods: Automatic, Manual, Combined, or Workflow. You can start recognition from other applications, using Direct OCR and can also schedule processing to run at a later time.

Processing Documents

Processing methods

Using OmniPage, you can choose from the following processing methods:

Automatic



A fast and easy way to process documents is to let OmniPage do it automatically for you. Select settings in the Options dialog box and in the OmniPage Toolbox drop-down lists and then click *Start*. It will take each page through the whole process from beginning to end, when possible running in parallel. It will typically auto-zone the pages.

Manual



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

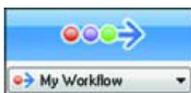
1. Use button one to get a set of images.
2. Manually zone pages where you want to process only part of the page or if you want to give precise zoning instructions. Use ignore backgrounds or zones to exclude areas from processing. Use process backgrounds or zones to specify areas to be autozoned.
3. Use button two to have the pages recognized.
4. Do proofing and editing as desired.
5. Use button three to save your results.

The default for manual processing is to have all entered pages automatically selected. This way you can have all new pages recognized by a single mouse click. You can remove this default in the Process panel of the Options dialog box.

Combined

You can process a document automatically and view results in the Text Editor. If most pages are in order, but a few have not turned out as expected, you can switch to manual processing to adjust settings and re-recognize just those problem pages. Alternatively, you can acquire images with manual processing, draw zones on some or all of them, and then send all pages to automatic processing by pressing the Start button and choosing to process existing pages.

Workflow



A workflow consists of a series of steps and their settings. Typically it will include a recognition step, but it does not have to. It does not have to conform to the 1-2-3 pattern of traditional processing. Workflows are listed

in the Workflow drop-down list – sample workflows plus any you create. Workflows allow you to handle recurring tasks more efficiently, because all the steps and their settings are pre-defined. You can choose to place the OmniPage Agent icon on your taskbar. Its shortcut menu lists your workflows. Click a workflow to launch OmniPage and have it run. Let the Workflow Assistant guide you in creating new workflows. It provides a choice of steps and the settings they need. Click Next after each step to add another one. You can use the Assistant just to get more guidance when doing automatic processing.

Processing automatically

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

To process your document automatically:

1. Make sure **1-2-3** is selected in the **Workflow** drop-down list.
2. Select the desired **Get Page** setting in the drop-down list.

You define the document source, which can be from image files or from a scanner.

3. Select a setting from the **Layout Description** drop-down list.

This guides the program in auto-zoning the pages. You describe the incoming pages or specify a zone template file.

4. Select a setting from the **Export Results** drop-down list.

You can save recognized pages to file, copy them to Clipboard, send them as mail attachments or direct them to other targets.

- Save the document as an OmniPage Document file from the File menu or the Standard toolbar.
5. Choose **Options...** in the Tools menu and check that settings are appropriate for your document.

You can, for instance, specify recognition languages and whether you want to proofread the document or not. You can also load an image enhancement template, a training file or a settings file.

6. Click the Start button or choose **Workflows / Start** in the Process menu with **1-2-3** selected.

Each page of the document is processed and finished one after the other. The program may perform tasks simultaneously, for instance it may start loading and recognizing a new page as you proofread the previous page.

7. The Workflow Assistant is unavailable during automatic processing. You can, however, press the **Stop** button and finish processing manually.
- You may reprocess all pages if an unsuitable setting caused poor results on all pages. An example is incorrect language choice, resulting in almost all words marked suspect during proofing. To do this, press the Start/Stop button, change settings, click Start again and select **Re-process all existing pages without adding new ones** which lets you perform re-recognition without having to scan or load or rezone all the images again.

Processing manually

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

To process your document manually:

1. Select **1-2-3** in the **Workflow** drop-down list.
2. Click in the Standard toolbar or Options... in the Tools menu to check or make settings in the Options dialog box.
3. Select the desired setting in the **Get Page** drop-down list.

You define the document source, which can be from image files or from a scanner. Access the scanner_settings dialog box and make settings as desired.

4. Click the **Get Page** button.

This either brings up a dialog box allowing you to name images files, or initiates scanning. The result is a document with one or more thumbnails in the Thumbnails panel and one page image displayed in the Page Image panel.

5. Examine the image display – enhance the images if desired.
6. Now you can manually draw zones on one or more images and assign their properties.

Status bar buttons let you move to other pages. Any image without zones will be auto-zoned when recognition is requested. For more detail see Drawing zones manually and Zones, backgrounds and auto-zoning.

7. Select a setting in the **Layout Description** drop-down list. You describe the layout of the incoming pages. This setting has an influence if auto-zoning runs on any pages. You can also select a template to have its zones placed on the current page.

8. Click the **Perform OCR** button to have the current page recognized. To have selected pages recognized, make a multiple selection in the Document Manager or in Thumbnails and then click the **Perform OCR** button.
9. If you requested proofing, the OCR Proofreader dialog box displays suspect words one after the other from the recognized page(s).

You can proof and edit the recognized text.

10. To process and transfer zone changes to the Text Editor immediately click . This enables on-the fly processing of zoning changes made on recognized pages.
10. Continue loading pages, performing OCR, editing and proofing as desired.
11. Select a setting in the **Export Results** drop-down list.

You can save recognized pages (all, selected or current) to file, copy to Clipboard, send as mail or send to other targets. You can save pages more than once.

- Save the document as an OmniPage Document file from the File menu or the Standard toolbar.
12. Click the **Export Results** button.

Defining the source of page images

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

Input from image files

You can create image files from your own scanner, or receive them by e-mail or as fax files. OmniPage 16 can open a wide range of image file types. Select *Load Files* in the Get Pages drop-down list. Files are specified in the Load Files dialog box. This appears when you start automatic processing. In manual processing, click the Get Page button or use the Process menu. The lower part of the dialog box provides advanced settings, and can be shown or hidden. The minimum width or height for an image file is 16 by 16 pixels; the maximum is 8400 pixels (71cm or 28 inches at the resolution 201 to 600 dpi).

Input from digital camera

You can bring digital camera photos of documents for recognition into OmniPage. First, make sure that your device driver is installed properly. Then connect the camera and download images. Click Load Digital Camera Files in the Get Page drop-down list. If you use this, 3D Deskew,

resolution enhancement and straightening text lines are automatically performed on images. To acquire digital camera photos containing text from Direct OCR or PaperPort, mark the Load as digital camera image checkbox. The above mentioned automatic enhancements will apply.

Input from scanner

You must have a functioning, supported scanner correctly installed with OmniPage 16. You have a choice of scanning modes. In making your choice, there are two main considerations:

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



Scan black and white

Select this to scan in black-and-white. Black-and-white images can be scanned and handled quicker than others and occupy less disk space.



Scan grayscale

Select this to use grayscale scanning. For best OCR accuracy, use this for pages with varying or low contrast (not much difference between light and dark) and with text on colored or shaded backgrounds.



Scan color

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

Brightness and contrast

Good brightness and contrast settings play an important role in OCR accuracy. Set these in the Scanner panel of the Options dialog box or in your scanner's interface. After loading an image, check its appearance. If characters are thick and touching, lighten the brightness. If characters are thin and broken, darken it. Then rescan the page.

If your scanning results are still not satisfactory, open the scanned image in the Image Enhancement window to edit it using a range of different tools.

Scanning with an ADF

The best way to scan multi-page documents is with an Automatic Document Feeder (ADF). Simply load pages in the correct order into the ADF. You can scan double-sided documents with an ADF. A duplex scanner will manage this automatically.

Scanning without an ADF

Using OmniPage's scanner interface, you can scan multi-page documents efficiently from a flatbed scanner, even without an ADF. Select *Automatically scan pages* in the Scanner panel of the Options dialog box, and define a pause value in seconds. Then the scanner will make scanning passes automatically, pausing between each scan by the defined number of seconds, giving you time to place the next page.

Document to document conversion

In OmniPage Professional 16 you can open not only image files, but also documents created in wordprocessing and similar applications. Supported file types include .doc, .xls, .ppt, .rtf, .wpd and others. Click the Load Files button in the OmniPage Toolbox or select the Load Files command under Get Page, in the File menu. In the Load Files dialog box, choose Documents. When you are finished, you can choose from a wide variety of document file types for saving.

Describing the layout of the document

Before starting recognition you are requested to describe the layout of the incoming pages to assist the auto-zoning process. When you do automatic processing, auto-zoning always runs unless you specify a template that does not contain a process zone or background. When you do manual processing, auto-zoning sometimes runs

Here are your input description choices:

Automatic

Choose this to let the program make all auto-zoning decisions. It decides whether text is in columns or not, whether an item is a graphic or text to be recognized and whether to place tables or not.

Single column, no table

Choose this setting if your pages contain only one column of text and no table. Business letters or pages from a book are normally like this.

Multiple columns, no table

Choose this if some of your pages contain text in columns and you want this decolumnized or kept in separate columns, similar to the original layout.

Single column with table

Choose this if your page contains only one column of text and a table.

Spreadsheet

Choose this if your whole page consists of a table which you want to export to a spreadsheet program, or have treated as single table.

Form

Choose this if your whole page consists of a form and you want form elements auto-recognized. After recognition, you can modify form element properties, create new ones, or edit form layout.

Legal pleading

Choose this to recognize legal documents. Legal headers are detected and removed. Choose to have pleading numbers retained or dropped.

Custom

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

Template

Choose a zone template file if you wish to have its background value, zones and properties applied to all acquired pages from now on. The template zones are also applied to the current page, replacing any existing zones. If auto-zoning yielded unexpected recognition results, use manual processing to rezone individual pages and re-recognize them.

User dictionaries

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

Starting a user dictionary

Click **Add** in the OCR Proofreader dialog box with no user dictionary loaded or open the User Dictionary Files dialog box from the **Tools** menu and click **New**.

Loading or unloading a user dictionary

Do this from the **OCR panel** of the **Options** dialog box or from the **User Dictionary Files** dialog box.

Editing or removing a user dictionary

Add words by loading a user dictionary and then clicking Add in the OCR Proofreader dialog box. You can add and delete words by clicking Edit in the User Dictionary Files dialog box. You can also import words from OmniPage user dictionaries (*.ud). While editing a user dictionary,

you can import a word list from a plain text file to add words to the dictionary quickly. Each word must be on a separate line with no punctuation at the start or end of the word. The Remove button lets you remove the selected user dictionary from the list.

To embed a user dictionary in an OmniPage Document, load your input file, choose Tools > User Dictionary; select the user dictionary you want to use, click Embed, and name it. Then save to the file type OmniPage Document.

Text and image editing

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

Editing character attributes

In all views except Plain Text view, you can change the font type, size and attributes (bold, italic, underlined) for selected text.

Editing paragraph attributes

In all views except Plain Text view, you can change the alignment of selected paragraphs and apply bulleting to paragraphs

Paragraph styles

Paragraph styles are auto-detected during recognition. A list of styles is built up and presented in a selection box on the left of the Formatting toolbar. Use this to assign a style to selected paragraphs.

Graphics

You can edit the contents of a selected graphic if you have an image editor in your computer. Click Edit Picture With in the Format menu. Here you can choose to use the image editor associated with BMP files in your Windows system, and load the graphic. Alternatively, you can use the Choose Program... item to select another program. This will replace the Default Image Editor item. Edit the graphic, then close the editor to have it re-embedded in the Text Editor. Do not change the graphic's size, resolution or type, because this will prevent the re-embedding. You can also edit images before recognition using the Image Enhancement tools.

Tables

Tables are displayed in the Text Editor in grids. Move the cursor into a table area. It changes appearance, allowing you to move gridlines. You can also use the Text Editor's rulers to modify a table. Modify the placement of text in table cells with the alignment buttons in the Formatting toolbar and the tab controls in the ruler.

Hyperlinks

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

Editing in True Page

Page elements are contained in text boxes, table boxes and picture boxes. These usually correspond to text, table and graphic zones in the image. Click inside an element to see the box border; they have the same coloring as the corresponding zones.

Saving and exporting

Once you have acquired at least one image for a document, you can export the image(s) to file. Once you have recognized at least one page, you can export recognition results – a single page, selected pages or the whole document – to a target application by saving to file, copying to Clipboard or sending to a mailing application. Saving as an OmniPage Document is always possible. OmniPage provides comprehensive support for Office 2007 applications and formats.

A document remains in OmniPage after export. This allows you to save copy or send its pages repeatedly, for example with different formatting levels, using different file types, names or locations. You can also add or re-recognize pages or modify the recognized text. With automatic processing and in Batch Manager jobs, you specify where to save first before processing starts.

A workflow may contain one or more saving steps, even to different targets (for instance, to file and to mail). A Batch Manager job must contain at least one saving step.

If you want to work with your document again in OmniPage in a later session, save it as an OmniPage Document. This is a special output file type. It saves the original images together with the recognition results, settings and training. Exporting is done through button 3 on the OmniPage

Toolbox. It lists available export targets. Some appear only if access to the target is detected on your computer. Select the desired target then click the Export Results button to begin export. You can also perform exporting through the Process menu.

Saving original images

You can save original images to disk in a wide variety of file types with or without image enhancement (using the Image Enhancement Tools).

1. Choose Save to File in the Export Results drop-down list. In the dialog box that appears, select Image under Save as.
2. Choose a folder location and a file type. Type in a file name.
3. Select to save the selected zone image(s) only, the current page image, selected page images or all images in the document. For multiple zones or multiple pages, you can have all images in a single multi-page image file, providing you set TIFF, MAX, DCX, JB2 or Image-only PDF as file

type. Otherwise each image is placed in a separate file. OmniPage adds numerical suffixes to the file name you provide, to generate unique file names.

4. Click Converter Options... if you want to specify a saving mode (black-and-white, grayscale, color or 'As is'), a maximum resolution and other settings. For TIFF files, you specify the compression method here.

5. Click OK to save the image(s) as specified. Zones and recognized text are not saved with the file.

Saving recognition results

You can save recognized pages to disk in a wide variety of file types.

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

2. The Save to File dialog box appears. Select **Text** under Save as.

3. Select a folder location and a file type for your document. Select a page range, file options, naming options and a formatting level for the document.

4. Type in a file name. Click **Converter Options**, if you want to specify precise settings for the export.

5. Click **OK**. The document is saved to disk as specified. If *View Result* is selected, the exported file will appear in its target application; that is the one associated with the selected file type in your Windows system or in the advanced saving options for your selected file type converter.

Selecting a formatting level

The formatting level for export is defined at export time, in the saving dialog box (Save to File, Copy to Clipboard, Send in Mail or other dialog box). Three of the levels correspond to the format views of the same name in the Text Editor. However, the level to be applied for saving is independent of the formatting view displayed in the Text Editor. When exporting to file or mail, first specify a file type. This determines which formatting levels are available.

The formatting levels are:

Plain Text

This exports plain decolumnized left-aligned text in a single font and font size. When exporting to Text or Unicode file types, graphics and tables are not supported. You can export plain text to nearly all file types and target applications; in these cases graphics, tables and bullets can be retained.

Formatted Text

This exports decolumnized text with font and paragraph styling, along with graphics and tables. This is available for nearly all file types.

Flowing Page

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

True Page

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

Spreadsheet

This exports recognition results in tabular form, suitable for use in spreadsheet applications. This places each document page onto a separate worksheet.

When exporting to Microsoft Excel, 'Spreadsheet' is good for saving whole-page tables. Prefer 'Formatted Text' if your document contains smaller tables: each table will be placed on a separate worksheet with non-table parts placed in an index worksheet with hyperlinks to each relevant worksheet

Selecting converter options

Click the Converter Options... button in a saving dialog box to have precise control over the export. This brings up a dialog box with the name of the converter associated with the current file type. It presents a series of options tailored to this file type. First, confirm or change the formatting level, because this influences which other options are presented. Select options as desired.

Using multiple converters

Multiple converters allow you to export to two or more file types in one export step. Choose Multiple in the saving dialog box: To make your own multiple converter, open the Save Preferences dialog box from the Tools menu. Choose the heading Multiple converters. Select a converter and click Create from... . This will make a copy of the selected converter that you can freely modify without overwriting the original one.

The new converter appears in the list. Select it and click Options... to specify its settings. You receive a list of all text converters, followed by all image converters. Checkmark the desired ones. Optionally specify sub-folder paths for each file type.

You can save pages with different formatting levels or file options to the different file types, as defined in their simple converters. A few saving operations cannot be done with multiple converters. These are:

Saving OmniPage Documents

Use a workflow with two saving steps, or perform two separate saves.

Saving to two targets

For instance, you cannot use a multiple converter to save a document to file and also send it in mail. Use a workflow with two saving steps, or perform two separate saves.

Saving different page ranges

You cannot save different page ranges to different file types, because only one set of selected pages can exist at saving time. For the same reason, a single workflow cannot be used either. Perform two separate saves or use two workflows.

Saving to PDF

You have five choices when saving to Portable Document Format (PDF) files. The first four are presented as Text converters, the last one is listed among the Image converters.

PDF (Normal):

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

PDF Edited:

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

PDF Searchable Image (formerly PDF Image on Text):

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

Quick Start Guide

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

Loading and recognizing sample image files

You will find sample image files in the program folder, both single-page and multi-page files. First try reading these files using the procedure presented below, except for the references to a

scanner. The results provide you with a benchmark of the recognition quality you should expect from your own files of comparable quality.

Next, try scanning a page from your scanner.

Scanning and recognizing a single page

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

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

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

	What you do	What happens
1.	Setup your scanner using the Scanner Wizard, if this is not already done.	Configures OmniPage to work with your scanner.
2.	Select Start > All Programs > ScanSoft OmniPage 16 > OmniPage 16	Opens OmniPage on your computer.
3.	Place the document correctly in your scanner.	
4.	From the Get Page drop-down list, select a scan option for your document: black-and-white, grayscale or color.	Allows you to determine how pictures or colored texts and backgrounds will look in the exported document. Scan Color is available only if you have a color scanner.
5.	From the Layout Description drop down list, check Automatic is selected. For a wide range of documents, this is the best choice.	Configures the program to place zones on the page and decide their properties automatically.
6.	From the Export Results drop-down list, check that Save to File is selected.	This means you will be able to name your export file after you have proofed the document.
7.	Make sure 1-2-3 is selected in the Workflow drop-down list. Click the Start button.	OmniPage will start to scan in your document. A thumbnail appears with a progress indicator. The OCR Proofreader appears.
8.	Use the <u>OCR Proofreader</u> to modify words that the program suspects have not been recognized correctly.	The OCR Proofreader operates like a spelling checker in a word processing program, but with added OCR-specific features. It removes markings from words you proof.
9.	Click in the Text Editor. Select Text Editor views one after another, to see how the page appears in each view.	Each Text Editor view defines a formatting level. This guides you which level to choose at saving time.

10.	Click <i>Resume</i> to restart proofing. When the message OCR Proofreading is complete appears, click on OK.	This ends the OCR Proofreader process. The Save to File dialog box will appear.
11.	Choose a file name, file type, path and formatting level to save your recognized document. Click on OK.	By default, View result is enabled, so your document will be automatically opened in the word processing program associated with the file type that you selected.
12.	Inspect the document in your word processing program.	You have successfully used OmniPage 16 to recognize your document and open it in your target application!