

1 OVERVIEW

Welcome to Xerox Imaging Systems' **TextBridge™ OCR for Windows**. OCR stands for **optical character recognition**, the capability to read an image and recognize and output text from it. Images can come from scanners, fax modems, or other sources.

TextBridge is a suite of applications that combines Xerox Imaging Systems' industry-leading OCR with simple Microsoft Windows interfaces.

Figure 1-1 shows the TextBridge Main dialog, which you can access as a stand-alone program or embedded in other text applications.



Figure 1-1. Main dialog

To work with imaging and fax applications, TextBridge provides an **OCR Printer**, which is selected through the application's Print Setup command. Then, using the Print command of the application, you can "print" an image to the OCR Printer, and save recognized text to a file in the text format of choice (for example, your word processor format).

TextBridge applications can convert recognized text to a number of word processing, spreadsheet, database, and other text formats.

TextBridge runs on Windows-compatible personal computers with an 80386 (or more powerful) microprocessor and at least four megabytes of system memory (eight megabytes is recommended).

The program runs on DOS Version 5.0 (or later) and Microsoft Windows, Version 3.1 (or later) in **enhanced mode** only.

TextBridge will also run under IBM's **OS/2®** operating system (version 2.0).

WHAT IS TEXTBRIDGE OCR?

TextBridge OCR is software that turns image data into usable text files on your PC.

With TextBridge OCR, you can access the valuable data locked inside paper documents, as well as on-line faxes and page images from other sources.

TextBridge recognizes and converts scanned and on-line page images to text files that you can open, edit, reformat, republish, and otherwise apply (Figure 1–2).

TextBridge supports a number of popular desktop scanners and converts recognized text to many popular **text formats**.

Use your scanner to input hard copy documents to TextBridge, which takes the scanned images, performs OCR, converts the recognized text to the format of your choice, and stores it as a PC file.

Alternatively, use TextBridge to recognize and convert on-line **page images** stored in **TIFF** (Tag Image File Format). These images can come from **fax modems** or other sources.

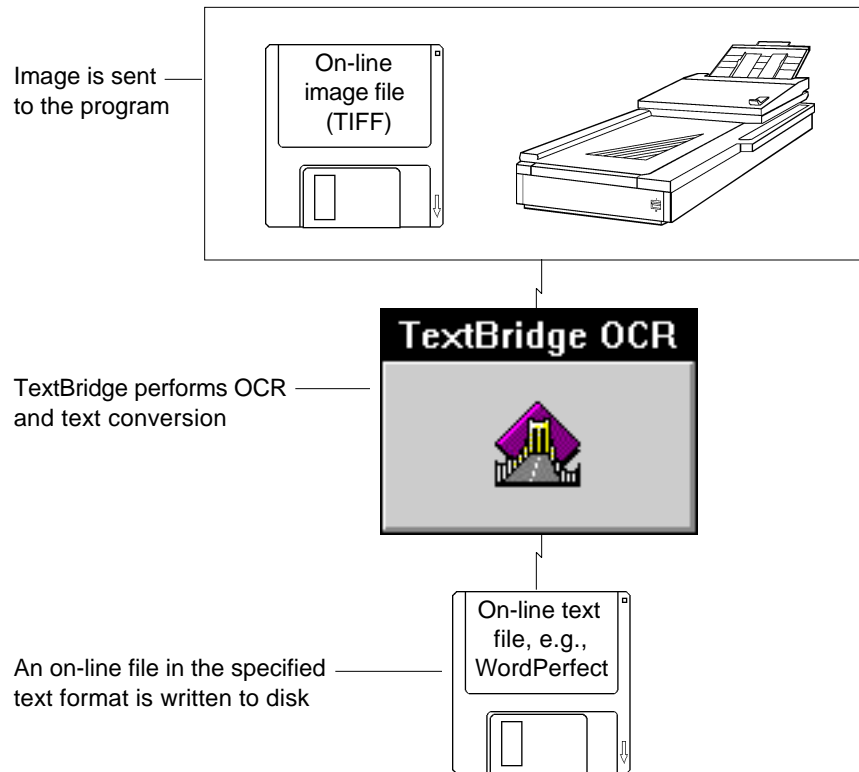


Figure 1-2. TextBridge OCR process

TextBridge provides an easy-to-use interface and a powerful set of built-in capabilities. For example, the **preview** tool lets you view and zone pages before OCR. The **verifier** lets you interact with the OCR software to achieve the highest possible accuracy.

TextBridge incorporates Xerox Imaging Systems' industry-leading document recognition software, and includes a number of technologies developed by Xerox Palo Alto Research Center (PARC), where modern computer interfaces originated.

Consequently, TextBridge provides the most accurate OCR and format retention results on the widest range of documents:

- documents with point sizes ranging from 6-point to 72-point type in practically any typeface
 - + TextBridge can reliably recognize type smaller than 8-point only from images scanned at 400 dots per inch resolution.
- documents printed on typewriters, phototype-setters, and impact, ink-jet, and laser printers
- photocopied, degraded, or dirty documents
- documents with single- or multiple-column layouts
- documents containing halftones
- on-line single- or multiple-page TIFF images from fax modems and other sources
- hard-copy faxes
- documents composed in English, French, Italian, German, or Spanish
 - + TextBridge versions shipped in international markets can recognize an even greater number of languages.

SUPPORTED TEXT FORMATS

TextBridge can convert recognized text to a number of output formats. With some formats, TextBridge supports multiple versions, as shown in Table 1–1.

Note that this list is subject to change. Refer to the on-line *TextBridge Release Notes* for the latest information.

Table 1–1. Supported Text Formats

Application	File Extension
Ami Pro 3.0	.sam
Ami Pro 2.0	.sam
ASCII Standard	.txt
ASCII Smart	.txt
ASCII Stripped	.txt
dBase IV	.dbf
DCA/RFT	.rft
DisplayWrite 5	.rft
FrameMaker	.mif
Interleaf	.wps
Lotus 1-2-3	.wk1
Excel for the Macintosh	.xls
Excel 3.0	.xls
Excel 4.0	.xls
RTF (Rich Text Format)	.rtf
RTF for the Macintosh	.rtf
Multimate Advantage	.doc
PostScript	.ps
Professional Write 2.0	.doc
Professional Write 2.2	.doc
Samna Word IV+	.sam
Windows Write	.wri
Word for Windows 2.0	.rtf
WordPerfect 4.2	.wpf
WordPerfect 5.1	.wpf
WordStar	.doc

SUPPORTED SCANNERS

Using built-in **ISIS** (Image and Scanner Interface Standard) drivers from Pixel Translations Inc., TextBridge works with most scanners in the PC market:

Apple OneScanner
Canon IX-12
Complete PC scanner
Ficus LEOScan 610
Ficus LEOScan 1210
Envision 6000
Envision 6100
Envision 8100
Epson ES300C (GT-4000 outside U.S.)
Epson ES600C (GT-6500 outside U.S.)
Epson ES800C (GT-8000 outside U.S.)
Fujitsu ScanPartner 10
Fujitsu M3096G
Fujitsu M3097G
HP ScanJet
HP ScanJet Plus
HP IIc
HP IIp
HP IIcx
Microtek MS-II
Microtek ScanMaker II
Microtek 600Z
Panasonic FX-RS307
Relisys Aries 1201
Tamarack 6000c
UMAX UC-630 scanner with GSII-PC card
XIS Datacopy GSplus
XIS Datacopy 730GS

In addition, TextBridge supports the **TWAIN** standard.

Thus, TextBridge works with any fully TWAIN-compliant scanner or other device that connects to a PC and produces binary (black-and-white) images in a supported size and resolution.

For example, a number of hand scanners, such as the Logitech™ ScanMan® and Artec WalkScan™, are provided with TWAIN source drivers.

In addition, some of the scanners that TextBridge supports through ISIS also are provided with TWAIN source drivers—for example, the Ficus LeoScan™.

Note The list of scanners that work with TextBridge is always growing. With TWAIN support, and the growing list of scanner manufacturers who are providing TWAIN drivers, TextBridge will work with many scanners. Refer to the on-line *TextBridge Release Notes* for the latest information.

WHAT COMES WITH TEXTBRIDGE

TextBridge is provided on several 3.5-inch high-density diskettes. The diskettes include software programs and libraries, ISIS drivers, **language packs**, and on-line help.

The TextBridge product also includes this user's guide and a quick-reference card, as well as a software registration card.

Note Be sure to fill out the software registration card, as it entitles you to technical support, and assures that you are kept up to date on new software releases.

If any piece is missing from your TextBridge package, call your authorized Xerox Imaging Systems' dealer or reseller. If you are unable to solve the problem, you can call Xerox Imaging Systems directly.

For information about contacting Xerox Imaging Systems, refer to the "Customer Support" section in the Preface of this user's guide.

SYSTEM REQUIREMENTS

To install and run TextBridge, your Windows-compatible PC must be equipped with the following:

- an Intel (or compatible) 80386, 80486, or Pentium™ microprocessor
- Microsoft Windows™, version 3.1 or later running in **enhanced mode** only
- Microsoft Disk Operating System (DOS)™, version 5.0 or later
- four megabytes (4Mb) of random access memory (RAM); eight megabytes (8Mb) is recommended
- 8Mb to 16Mb of **permanent virtual memory**; 16Mb is recommended
 - + For information about configuring your system with permanent virtual memory, refer to the *Microsoft® Windows™ Version 3.1 User's Guide* published by Microsoft Corporation.
- a hard disk with a minimum of 4Mb of free space in which to install TextBridge; the 4Mb minimum disk space requirement enables installation of all TextBridge application software, one ISIS scanner driver, and one language pack. Please allow 700K for each additional language pack you intend to install.

Note TextBridge runs under IBM's OS/2 operating system, Version 2.0 and later, which operates many Windows 3.1 programs. On OS/2 systems, TextBridge requires 16Mb of RAM.

ON-LINE HELP

TextBridge is designed to be easy to learn and use. However, if you need assistance, TextBridge provides a complete **hypertext**-based on-line Help system.

In any of TextBridge's screens, you can select Help and display a Help window about that particular screen (Figure 1–3).

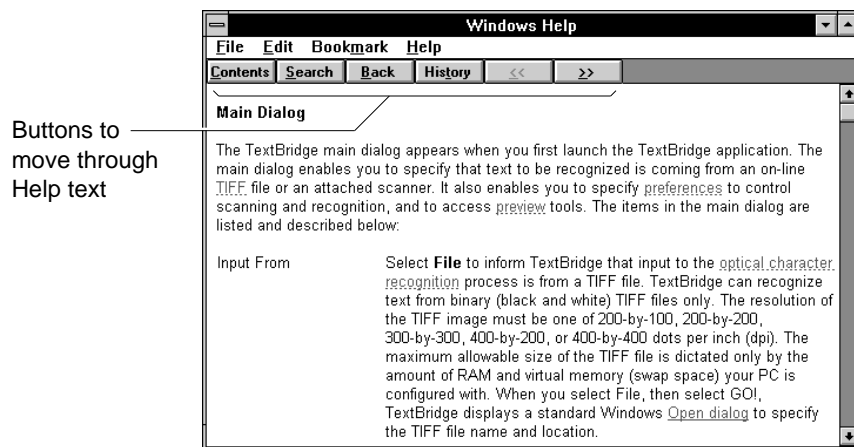


Figure 1–3. TextBridge on-line Help

By selecting the Contents button near the top of the window, you can display the top-level index for TextBridge Help. You can move through the TextBridge Help system in a number of ways:

- by selecting a topic from the Contents list
- by choosing a jump
- by searching for a topic
- by browsing forward or backward
- by backtracking

In any Help window, you can pull down the Help menu and select the “How to Use Help” topic for complete information about using Help tools.

GETTING STARTED

Before you can use TextBridge, you must install it on your PC as directed in Chapter 2.

- + If you run into any problems installing TextBridge, refer to Appendix A of this user's guide, which provides troubleshooting tips.

After you successfully install TextBridge, you can refer to the *Quick Reference Card* for a quick-start procedure.

Alternatively, for more detailed step-by-step operating procedures, refer to Chapter 3 of this guide.

After you become familiar with the basic operation of TextBridge, refer to Chapter 4, "Tips and Techniques." This chapter provides advice for getting the most out of TextBridge.