

Xerox TextBridge ActiveX Control

User's Guide

IMAGE  BASIC

Diamond Head Software, Inc.
1217 Digital Drive Ste. 125
Richardson, Texas 75081
(972) 479-9205

Copyright Notices

Under the copyright laws, neither the documentation nor the software may be copied, photocopied, reproduced, translated, or reduced to any electronic medium or machine-readable form, in whole or in part, without the prior written consent of Diamond Head Software, Inc., except in the manner described in the documentation.

This software product contains proprietary software components developed by a number of different software companies, referred herein as "Third Party Licensors". This documentation and the software are protected by one or more of the following copyright notices:

Portions of this product, © 1994, 1995, 1996, 1997 Diamond Head Software, Inc. All rights reserved.

Portions of this product, © 1994, 1995, 1996 Xerox Imaging Systems, Inc. All rights reserved.

Company and product names mentioned in this documentation are trademarks or registered trademarks of their respective companies. Lotus and Lotus Notes are registered trademarks of Lotus Development Corporation. Windows is a trademark and Microsoft is a registered trademark of Microsoft Corporation.

DIAMOND HEAD SOFTWARE INC. AND ITS THIRD PARTY LICENSORS MAKE NO WARRANTIES, EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, REGARDING THE SOFTWARE. DIAMOND HEAD SOFTWARE, INC. AND ITS THIRD PARTY LICENSORS DO NOT WARRANT, GUARANTEE OR MAKE ANY REPRESENTATIONS REGARDING THE USE OR THE RESULTS OF THE USE OF THE SOFTWARE IN TERMS OF ITS CORRECTNESS, ACCURACY, RELIABILITY, CURRENTNESS OR OTHERWISE. THE ENTIRE RISK AS TO THE RESULTS AND PERFORMANCE OF THE SOFTWARE IS ASSUMED BY YOU. THE EXCLUSION OF IMPLIED WARRANTIES IS NOT PERMITTED BY SOME JURISDICTIONS. THE ABOVE EXCLUSION MAY NOT APPLY TO YOU.

IN NO EVENT WILL DIAMOND HEAD SOFTWARE INC. OR ITS THIRD PARTY LICENSORS AND/OR THEIR DIRECTORS, OFFICERS, EMPLOYEES OR AGENTS BE LIABLE TO YOU FOR ANY CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES (INCLUDING DAMAGES FOR LOSS OF BUSINESS PROFITS, BUSINESS INTERRUPTION, LOSS OF BUSINESS INFORMATION, AND THE LIKE) ARISING OUT OF THE USE OR INABILITY TO USE THE SOFTWARE EVEN IF DIAMOND HEAD SOFTWARE INC. OR ITS THIRD PARTY LICENSORS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. BECAUSE SOME JURISDICTIONS DO NOT ALLOW THE EXCLUSION OR LIMITATION OF LIABILITY FOR CONSEQUENTIAL OR INCIDENTAL DAMAGES, THE ABOVE LIMITATIONS MAY NOT APPLY TO YOU. Diamond Head Software Inc.'s and its Third Party Licensors' liability to you for actual damages from any cause whatsoever, and regardless of the form of the action (whether in contract, tort (including negligence), product liability or otherwise), will be limited to

\$50.

Contents

Chapter 1 : Introduction to TextBridge	1
Linking of ImageBASIC Controls.....	1
Licensing Configuration and Verification.....	2
Chapter 2 : Using the TextBridge Control	5
Optical Character Recognition Using TextBridge.....	5
Setting Source and Destination.....	6
Output Formats.....	9
Starting a Recognition Attempt.....	10
Optimization of the Recognition Attempt.....	13
Language and Field Type Specification.....	15
Text Layout and Quality.....	21
Verification and Training.....	26
Verification.....	26
Chapter 3 : Reference	35
Reference to Properties, Methods and Events.....	35
Appendix A : Supported Formats and DLLs	71
Output Formats and Required Conversion Libraries.....	71
Index	73

Chapter 1 : Introduction to TextBridge

Linking of ImageBASIC Controls

Virtually all ImageBASIC controls can accept image data from other ImageBASIC controls. The process of designating where each ImageBASIC component gets its image data is referred to as linking the controls. With the exception of those controls that can directly access files or scanners, all ImageBASIC controls must be linked to another ImageBASIC control to get any image data.

Linking of controls is the primary method of moving an image through a series of processing steps. For example, an image may be originally captured through the PixScan control, passed to a TMSSequoia Display control for operator verification, optionally routed through a ScanFix control for enhancement, then to a TMSSequoia File control to be written to disk, and finally to a TextBridge control for OCR processing to generate indexing information for that file.

Creating the Link

Almost all of the ImageBASIC controls have a property named **ImageDataSource**. To create the data link between controls, this property must specify the ImageBASIC control that will be supplying image data. Any image that is received by the source control will also be sent to the linked control. For example, if the **ImageDataSource** property of a TextBridge control is set to a TMSSequoia Display control, each time a new image is loaded into the display window, the TextBridge control will receive that image.

Linking at Design Time

As each ImageBASIC control is added to a Form at design time, its **ImageDataSource** property is automatically set to an ImageBASIC control already on that Form. The assignment may be changed at design time by selecting from the drop-down list of available ImageBASIC components. This list is shown with the **ImageDataSource** property in the Properties Window or Object Inspector.

Linking at Runtime

Linking controls at runtime requires only one line of code that can be executed at any time. The source of image data can be changed during program execution by naming another ImageBASIC control in the **ImageDataSource** property, as shown here:

```
TextBridge1.ImageDataSource = TMSDispl.Link
```

Each time the image data flowing from one ImageBASIC control changes, the receiving control's **ImageDataChanged** event is triggered. From this event, any procedure that is to be performed on each image can be started. For example, each time an OCR control's **ImageDataChanged** event occurs, an OCR attempt could be started on the new image data.

Licensing Configuration and Verification

In order to run, each ImageBASIC control must be able to verify the presence of a valid license token. These license tokens are stored on either a hardware key (shipped with each toolkit purchase) or in a licensing database (the typical toolkit evaluation and runtime distribution format).

Using these tokens, licensing in ImageBASIC is based on enabling a set number of concurrent seats. Each license token allows a single PC to run any number of instances of a single control. For example, a single token for a Display control will allow one workstation to concurrently run multiple applications, each of which employs any number of Display controls.

A unique token is required for each different type of ImageBASIC component that you have licensed:

- There is a special type of token for the TMSSequoia Display control, another for the TextBridge control, another for the ScanFix control, and yet more token types for each additional control. The 16-bit and 32-bit versions of each control are also licensed separately.
- Each one of these licenses also comes in two varieties: *runtime* and *development*.

As suggested by the names, runtime licenses are necessary for an executable to function, and development licenses are necessary to develop an application using ImageBASIC.

A design time license will function as a runtime license, removing the need to add runtime tokens for application testing during development.

Where the Licenses are Kept

The ImageBASIC licensing server will find licenses stored in either one of two locations -- in a licensing database or on a hardware key:

- The hardware key is plugged into a parallel port on the computer using ImageBASIC and will be automatically found each time an ImageBASIC component is used.

Note: When using Windows NT, the parallel port must be configured using the Sentinel drivers that are installed with ImageBASIC in the HARDKEY subdirectory.

- The licensing database must be created on the site where it will be used and may not be moved from the location in which it is installed.

A file called IMGBASIC.INI must be in the Windows directory of each PC that is using an license database. This file contains an entry pointing to the database.

Creating a Licensing Database

Licensing databases can be installed in two ways:

- 1) If you are installing an evaluation copy of the ImageBASIC CD, run the registration utility that is accessible through the CD Navigator.
- 2) If you are installing a runtime database or are modifying an existing database, run a utility called the License Configuration Manager (LCM). The LCM is installed along with all ImageBASIC toolkit in the TOOLS subdirectory.

How the License Tokens are Used

As each application that uses ImageBASIC is initiated, or the control is loaded into the development environment, the ImageBASIC licensing server attempts to find the proper token for each component.

When the ImageBASIC controls are loaded into a development environment, the licensing server will always immediately attempt to verify that the controls are licensed.

When the controls are loaded at runtime, the licensing server can be instructed to delay the verification of available licenses. For this purpose, each control has a property named **Active**. The **Active** property can be changed to False only at design time. If the property is False when a control initializes at runtime, licensing will not be verified and the control's technology libraries will not be loaded. This will make the initialization of an application somewhat faster.

Before using any control that was loaded without full initialization, the **Active** property must be set to True by the application. At this time, the licensing server will find and lock the requisite token and the technology libraries will be loaded. If the proper token cannot be found **Active** will be set back to False and the control cannot be used.

Licensing Token Release

When an application that has locked one or more tokens terminates normally, the tokens are released and can be taken by another user if a network licensing database is being used. This is the process by which concurrent licensing for any number of seats may be enabled.

- If the application ends abnormally -- the user might reboot or a concurrently running Windows application might lock up -- then the release of the licenses is conditional on the network or disk operating system.
- If the licensing database has been installed locally, the locks are immediately released and will again be available when the application is started again.

Chapter 2 : Using the TextBridge Control

Optical Character Recognition Using TextBridge

The TextBridge control incorporates the powerful and flexible character recognition engine from Xerox Corp. The primary advantage of this recognition engine is its ability to learn and be trained to accurately recognize even difficult or distorted characters. The TextBridge engine is based on neural net technology that both incorporates user input and teaches itself to recognize new or difficult image text.

This chapter of the User's Guide introduces both basic recognition options and more advanced issues of specialized field types, user-defined dictionaries, and interactive verification.

The basic process of performing OCR using this control is as follows:

- 1) Specify an Image Source

The image source may be from file or from another ImageBASIC control and is specified through the **InputFrom** property.

- 2) Select Image Region(s) for Recognition

- 3) Select an output destination.

The recognition string output by the control may be sent either to the **Result** property for immediate reference or to a disk file, as specified through the **OutputTo** property.

- 4) Select the output format of the recognition string.

Supported output types range from plain ASCII text to text containing analysis markings to many popular word processor formats. The format is selected through the **OutputFormat** property. Remember that when creating output in any format other than the first six in the enumerated list, the output must be written directly to file.

- 5) Set processing options.

These options range from specifying the language of the document to the text layout of the text in that document. Refer to **Optimization of the Recognition Attempt** on page 13 for details.

- 6) Begin OCR on either the entire document page or on a region of the page by executing either the **OCRPage** or **OCRRegion** method.

Setting Source and Destination

The TextBridge control can accept image data for OCR from another ImageBASIC control, or it can read the image directly from file. Likewise, the output string of the recognition results may be written either to a property of the control or directly to file. The **InputFrom** and **OutputTo** properties specify the source and destination of the image and text, respectively.

Specify Image Source

The TextBridge control can accept image data from another ImageBASIC control or can read directly from file. Regardless of the specified source of image data, all other processing options and features are identical. The specification of the source of image data is made through the **InputFrom** property.

InputFrom is an enumerated property with the following valid options:

- 0 ImageDataSource
- 1 File

0--ImageDataSource causes the control to accept image data from the ImageBASIC control that is named in the **ImageDataSource** property.

1--File instructs the TextBridge control to read directly from an image file. When reading directly from file, the following property specifies the file to read:

- | | |
|---------------|--|
| InputFileName | A string property that specifies the fully qualified path and file name of the image file to read. If a multiple page image file is supplied, all pages in the file will be processed. |
| PageIndex | If a multiple page file is specified, this integer property specifies which page in the file to OCR. If set to 0 (zero), the default, all pages of the file will be processed. |

The following file formats may be read directly by the TextBridge control. In order to perform OCR on any other image file format, the file must be read by an ImageBASIC file control. The TextBridge control can then accept the image data from memory by linking it to the control that opened the file.

- TIFF Group 4
- TIFF Group 3
- TIFF Group 3 Modified
- TIFF Uncompressed
- TIFF Packed

Define Region(s) for Recognition

The TextBridge control can perform recognition of the entire image or of one or more regions within the image. The source of the image is specified in the **InputFrom** property -- image data is received either from another ImageBASIC control (specified in the **ImageDataSource** property) or from a file (specified in the **InputFileName** property).

*When the entire image page is to be recognized using the **OCRPage** method:*

- No regions should be defined within the TextBridge control. Excluding region definitions is not necessary when using the **OCRPage** method but can help avoid confusion.
- The **RegionSource** property can be set to any valid value, but the region coordinates from that source will be ignored.

*When a region or regions are to be recognized using the **OCRRegion** method:*

- TextBridge will first query the **RegionSource** property. If this specifies a valid source of region coordinates, that single region will be recognized.
- If no **RegionSource** is specified, then TextBridge will then query the region definition properties, described below. If one or more regions is defined, all defined regions will be recognized.
- Finally, if no **RegionSource** is available and if no region properties are set, the entire input image will be recognized.

In all cases, the OCR results will be output as specified in the **OutputTop** property.

Region Definition Properties and Methods

The TextBridge control can define one or more regions for recognition using the **OCRRegion** method. The following region definition properties specify the region coordinate in image pixels:

RegLeft

RegTop

RegRight

RegBottom

The following additional properties are available to select a single region for definition or review:

RegCount Reports the total number of regions that are allocated.

RegIndex	May be set to any integer from 1 (one) to RegCount to refresh the region definition properties for that region.
----------	--

The TextBridge control initializes with no regions defined. A new region may be allocated using the following method:

AddRegion	When executed, the RegCount property is updated and a new region is initialized with all region coordinates set to 0 (zero).
-----------	---

The region that is currently specified in the RegIndex property may be deleted using the following method:

DeleteRegion	
--------------	--

Select Output Destination

After TextBridge completes a recognition attempt the output string will be reported to one of two destinations, based on the value of the **OutputTo** property. This enumerated property has two valid options:

- 0 String
- 1 File

0--String populates the **Result** property with the output string. Under most circumstances, the output format of the results should be set to one of the ASCII text options (the first six in the enumerated **OutputFormat** property) when the OCR results are sent to the **Result** property.

1--File writes the OCR results string to a file. Several properties set parameters to the file writing process:

OutputFileName	A string property that specifies the fully qualified path and file name to the output file.
OutputFormat	An enumerated property that specifies the format of the output file. Options range from several configurations of ASCII text to popular word processing documents. Options for this property are detailed in the section 'Output Formats' on page 9.
OutputFileAppend	If True and the file named in the OutputFileName property already exists, the OCR results string will be appended to the end of the file. If False, the named file will be overwritten if it exists.

Note: Only plain text files can be appended. None of the word processor or spreadsheet formats can be appended by TextBridge.

If multiple regions are recognized in a single pass, the results from all regions will be concatenated in the output file. The output string is also reported as the return value from the **OCRPage** or **OCRRegion** method.

Output Formats

The following formats may be output by the TextBridge control. When recognition is complete, a text string will be returned by the **OCRPage** or **OCRRegion** method.

Note: If the **OutputFormat** property is set to any of the options that creates an output in other than options *1--None* through *5--Formatted ISO*, the TextBridge control must be allowed to write directly to file to save the data. Options for the **OutputFormat** property are as follows:

- 0 Default
- 1 None (0 byte file will be created)
- 2 Enhanced XDOC
- 3 Formatted ASCII Text
- 4 XDOC without Format Analysis
- 5 Formatted ISO
- 6 Forms ISO
- 7 Ami Pro 2.0
- 8 Ami Pro 3.0
- 12 ASCII Smart
- 13 ASCII Standard
- 14 ASCII Standard DOS
- 15 ASCII Stripped
- 16 dBase IV 1.0
- 17 DCA/RTF
- 18 Display Write
- 20 Excel for Macintosh
- 21 Excel 3.0
- 22 Excel 4.0
- 32 FrameMaker
- 33 Interleaf

34	Lotus 1-2-3
38	MultiMate Advantage II
43	Postscript
44	Professional Write 2.0
45	Professional Write 2.2
47	RTF (Rich Text Format)
48	RTF Macintosh
49	RTF MS Word 6.0
50	Samna Word IV
51	Windows Write
53	Word For Windows 2.x
54	Word For Windows 6.x
55	WordPerfect (Windows)
56	WordPerfect 4.2
59	WordPerfect 6.0
60	WordPerfect 6.1 (Windows)
64	WordStar 1.x (Windows)
68	HTML (2.0)
69	HTML (Netscape)
70	HTML (SoftQuad Editor)

Starting a Recognition Attempt

All of the options listed below should be set before initiating OCR.

- The core processes of selecting an image to OCR and setting the output are detailed above starting with "Setting Source and Destination" on page 6.
- Advanced options for improving OCR results based on characteristics of the input image are detailed starting with "Optimization of the Recognition Attempt" on page 13. For basic recognition, these options may all be left at their default values.
- During the OCR attempt, interactive verification and training of the recognition engine can occur, and these features are discussed starting with "Verification and Training" on page 26. The default settings for verification and training related properties disable these features.

After the image source and the output destination and format are specified, a new OCR attempt can begin. A new attempt is begun by executing one of the following methods:

OCRPage	Processes the entire input page without regard to its source or the definition of any regions within that source.
OCRRegion	Uses the image coordinates supplied through the RegionSource property or the region definition properties to process only the specified portion of the image.

Initiating Recognition

When image data is supplied to the TextBridge control, either from another ImageBASIC control through the **ImageDataSource** property or directly from a file specified in the **InputFileName** property, the entire page is available for processing.

- To perform OCR on the entire input page, execute the **OCRPage** method.
- To perform OCR on only a portion of the input page, execute the **OCRRegion** method. In order to OCR a region, the region(s) must be defined as discussed in the section titled 'Define Region(s) for Recognition' on page 7.

As of this writing, the primary control used to define a region for further processing is the TMSSequoia Display control. To define the region for an OCR attempt, link the TextBridge control to a TMSSequoia Display control and define the Working Region of the display control. The following sample code will create this link to the TMSSequoia Display control for both the **ImageDataSource** and the **RegionSource** of the TextBridge control and then perform OCR on that region, writing a file with the OCR results.

```

' create the image data links between ImageBASIC
' controls
TMSDispl.ImageDataSource = TMSFile1
TextBridgel.ImageDataSource = TMSDispl

' set the source of region coordinates for OCR
TextBridgel.RegionSource = TMSDispl

' select the ImageDataSource as the image input
TextBridgel.InputFrom = 0 ' ImageDataSource

' load a file; it will be displayed in the TMS
' Display control
TMSFile1.InputFileName = "c:\pending\10001.tif"
```

```

' specify a file as the text output
TextBridge1.OutputTo = 1 ' File

' specify output file type
TextBridge1.OutputFormat = 31 ' Windows Write file

' specify the file for output
TextBridge1.OutputFileName = "f:\ocr_out\10001.wri"

' specify the Working Region of the TMSSequoia
Display
' control; this is the portion of image data that
' will be supplied to the TextBridge control
TMSDispl.RegTop = 150
TMSDispl.RegBottom = 1500
TMSDispl.RegLeft = 250
TMSDispl.RegRight = 2000

' start OCR on the region; the output file will be
' written by TextBridge when recognition is
' complete
TextBridge1.OCRRegion

```

Events During Recognition

During the recognition attempt, the TextBridge engine can trigger certain events. These events can report errors or request information from the operator. Using these events, the developer can create a custom look and processing system based on individual requirements. The events that can occur as follows:

Completion	Occurs frequently throughout the processing performed by the control. Reports the current progress in the recognition process and allows canceling of the attempt.
Verify	<p>Will occur only if the Verify property is True.</p> <p>Occurs once for each character or group of characters which fall below the confidence level set in the ThresholdVerify property. During this event, the user can correct the recognition string.</p> <p>Corrections entered in this event will be retained for later use if the TrainFileName property is set to a valid path and file name and if the SaveTraining property is True. This same training file can be used in future recognition attempts if it is specified in the</p>

TrainFileName property and if the **LoadTraining** property is True.

Refer to "Verification and Training" on page 26 for details on the process and the use of the **Verify** event.

Error

Occurs each time an error internal to the TextBridge control is detected. This event reports the following parameters:

Number A long error code that identifies the error

Description Descriptive string of the error

SCode A composite long number indicating the severity of the error, the facility code, the origin of the error, and the native error code

Source Descriptive string of the source of the error

HelpFile Suggested help file name that should have a detailed explanation of the error

HelpContext Context ID of the appropriate topic in the help file named above

CancelDisplay If set to True during this event, the standard error dialog will not be displayed. If the *On Error* statement is used in the routine which began the recognition, the Visual Basic error message box will not be displayed.

Optimization of the Recognition Attempt

Optical Character Recognition is a complex and difficult task for any computer or program. In addition to the advantages available through its learning system and neural net technology, TextBridge will produce the most accurate recognition results when it is given as much information as possible before OCR begins.

For example, many images are of less than ideal clarity or resolution. In addition, many types of forms include regions of shading or inverse (white-on-black) text. Other documents are formatted in columns or irregular sections. The TextBridge control can adjust to and compensate for all of these discrepancies, at least to some degree. In order to do so, TextBridge must be instructed to expect images with these imperfections. Therefore, several properties are available to optimize the configuration of the recognition engine in an attempt to maximize accuracy and efficiency.

In addition to all of the configuration options shown in the following paragraphs, TextBridge can also be corrected and instructed during the recognition attempt. Please refer to 'Verification and Training' on page 26 for details on this feature.

Language and Field Type Specification

The TextBridge control allows the user to specify both dictionaries to use during OCR and a particular type of data (such as a phone number) to expect in a field. Two types of dictionaries, referred to as *lexicons*, are available.

The first type of lexicon is a language lexicon. A number of different language lexicons are included in the basic TextBridge installation. Selecting a language lexicon is discussed immediately below under the heading 'Language Lexicons' on page 15.

In addition, the user may define a separate lexicon for use in processing documents composed of unique or unusual words not found in the standard language lexicons. Refer to 'User-Defined Lexicon' on page 16 for details.

Several different field type specifications are available, including telephone numbers, dates, Roman numerals, and several others. Refer to 'Field Types' on page 17 for more information.

Lexicons, both language and user-defined, and field types are enabled and used together. See 'Enabling Lexicons and Field Types During OCR' on page 17 for information.

Language Lexicons

A number of different language files are available for use by the TextBridge control. These files enable the recognition engine to accurately process documents in many different languages. Using a language lexicon allows the recognition engine to assign a higher confidence level to any word that is found in the lexicon.

These language lexicons contain a standard dictionary for each language, but, as a rule, do not include technical or scientific terms. To optimize the recognition of documents containing relatively rare or specialized words, consider the use of a user-defined lexicon in addition to the standard language lexicon.

The specification of a language lexicon to use during OCR is made with the **Language** property. This property is set according to an enumerated list of these languages:

- 0 English
- 1 German
- 2 French
- 3 Spanish

- 4 Italian
- 5 Dutch
- 6 Norwegian
- 7 Finnish
- 8 Danish
- 9 Portuguese
- 10 Reserved
- 11 Russian
- 12 English & German

For details on enabling the use of language lexicons, refer to [Enabling Lexicons and Field Types During OCR](#) on page 17.

User-Defined Lexicons

When processing documents containing technical, scientific, or other exceptional words not found in standard dictionaries, the accuracy of the TextBridge control can be increased by the addition of a user-defined lexicon.

User-defined lexicons are text (ASCII) files which may contain any number of words. The formation of a new, user-defined lexicon is straightforward. Simply create an ASCII text file in any text editor or word processor that can save as text only.

- Each entry in a lexicon file must be on a separate line, separated from the next entry by a carriage return (ASCII code 0A hex) and line feed (0D hex) character.
- Each entry must consist of a single word without spaces.
- Entries are case-sensitive, so if an entry will appear with different capitalization, all of the possible forms should be entered in the lexicon.

A portion of a sample user-defined lexicon is shown below.

```
ImageBASIC
TextBridge
OCX
LexicalClass
VBX
Delphi
VB
```

photometric

After its creation, the user-defined dictionary will be used during recognition if it is specified in the **LexiconFileName** property and if the **LexicalMode** property enables the use of lexicons. For details on enabling the use of lexicons during recognition, see 'Enabling Lexicons and Field Types During OCR' on page 17.

Field Types

A number of specialized field types are defined within the TextBridge engine. When recognizing defined fields within a document, one or more of these field types may be specified to increase recognition accuracy.

The defined field types are as follows:

Date	Supports several date formats including <i>mm/dd/yy</i> and <i>month dd, yyyy</i>
Social Security Number	A standard U.S. social security number in the format <i>nnn-nn-nnnn</i>
Money	Recognizes both dollar and cent marks in addition to the monetary amount
Number	Expects only digits in this field; also recognizes comma and decimal point
Roman Numeral	Does not interpret the value of the number but does expect only known letters and does not attempt to identify the word in a lexicon
Telephone Number	Recognizes telephone numbers with and without area code

The expected field type is specified through the **LexicalClass** property. For details on the specification of field type and the use of lexicons during recognition, refer to 'Enabling Lexicons and Field Types During OCR' on page 17, below.

Enabling Lexicons and Field Types During OCR

Each time a new recognition attempt is begun, the TextBridge control checks the values in the lexicon-related properties:

Language

LexicalClass

LexicalMode

LexiconFileName

Specifying the Language of a Document

The enumerated property **Languages** specifies which of the installed language lexicons should be used during recognition. The available options are as follows:

- 0 English
- 1 German
- 2 French
- 3 Spanish
- 4 Italian
- 5 Dutch
- 6 Norwegian
- 7 Finnish
- 8 Danish
- 9 Portuguese
- 10 Reserved
- 11 Russian
- 12 English & German

Specifying a User-Defined Lexicon

The **LexiconFileName** property may contain the absolute path to a user-defined lexicon file. Refer to User-Defined Lexicon§ on page 16 for details on these files. A user-defined lexicon is a plain text file with one word on each line.

Activating Lexicon Use

The **LexicalMode** property specifies whether or not a lexicon is to be used during the next OCR attempt, and how closely the words returned by the recognition engine must match entries in the available lexicons. **LexicalMode** accepts values the following values:

- 0 No Lexical Processing
- 1 Preference
- 2 Absolute

0--No Lexical Processing

Indicates that the **Language** lexicon and the user-defined lexicon will not be accessed, even if one is specified in the **LexiconFileName**. Also, the setting of the **LexicalClass** property will be ignored and OCR will progress without any lexical help. Lexical processing may also be disabled in the **LexicalClass** property.

1--Preference

Takes advantage of the language lexicon, any user-defined lexicon, and any field types, as specified in, respectively, the **Language**, **LexiconFileName**, and **LexicalClass** properties. Unlike the *2--Absolute* option, the return string will reflect the operator input, even if it does not match the specified field type.

2--Absolute

If a recognized word, possibly as supplied by the operator **Verify** is True, does not match an entry in either the **Language** lexicon or one of the field types specified in **LexicalClass** or an entry in a user-defined lexicon, that word will be returned as a string of unrecognized characters.

For example, suppose the following:

- The **LexicalClass** property is set to *ID Number*, which recognizes U.S. Social Security numbers.
- TextBridge performs OCR on a field that contains the characters **445-83-6581**
- The engine recognizes this string as **445-83-65H1**

With **LexicalMode** set to *2--Absolute*, the following circumstances are possible:

- 1) If the **Verify** property is set to False, the Verification event will not occur, and the word will simply be returned as unrecognizable, or ~~~~~~
- 2) If the **Verify** property is True, the **Verify** event will occur, allowing the user to enter the correct string.

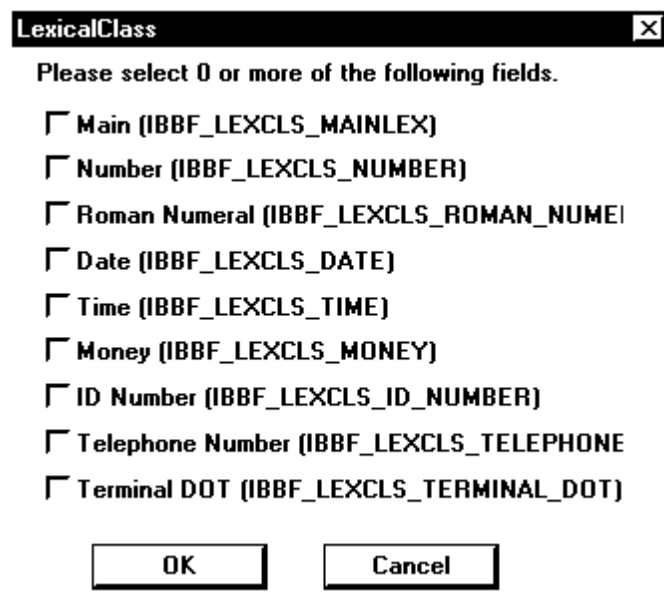
If the user confirms that the string is actually **445-83-6581**, the final output string will be that string.

If the user confirms the string as **445-83-65H1**, the final output string will consist of a series of the unrecognized character symbol, ~~~~~~; because a Social Security number cannot contain the letter **H**.

Specifying a Field Type

One or more field types may be selected through the **LexicalClass** property.

At design time, clicking on the property will display a dialog box listing all of the available field types. One or more of the check boxes next to each type may be marked. All of the marked field types will be used during recognition.



At runtime, the property may be set by adding together the constants for all of the field types that are being selected. The constant for each field type is listed below and is also listed in the design time dialog for this property. For example, to specify that the next region should be processed as either a time or date, set the property as shown here:

```
TxtBrg1.LexicalClass = IBBF_LEXCLS__DATE +  
                        IBBF_LEXCLS__TIME
```

To select no lexical class at runtime, set the property to 0 (zero).

The following options are available in the **LexicalClass** property:

Lexical Class	Constant
Main	IBBF_LEXCLS_MAINLEX
Number	IBBF_LEXCLS_NUMBER
Roman Numeral	IBBF_LEXCLS_ROMAN_NUMBER

Date	IBBF_LEXCLS_DATE
Time	IBBF_LEXCLS_TIME
Money	IBBF_LEXCLS_MONEY
ID Number	IBBF_LEXCLS_ID_NUMBER
Telephone Number	IBBF_LEXCLS_TELEPHONE
Terminal Dot	IBBF_LEXCLS_TERMINAL_DOT

Main causes recognition to proceed using the language lexicon and the user-defined lexicon.

Number recognizes the numerals 0-9, fractions, mathematical symbols and numbers ending *inst*, *nd*, *rd*, or *th*; for example, 1st, 2nd, etc.

Roman Numeral recognizes all characters that are valid in Roman numerals except single characters, three letter combinations beginning with *ml* or *m* (to avoid erroneous recognition of words like *mail*), and the combination *li* or *li* to avoid confusion with *ll*, *ll*, or *ll*.

Date recognizes any combination of a date, month, and year in numeral or with the month spelled or abbreviated. Both U.S. and European formats are recognized.

Time recognizes 12-hour and 24-hour times including any variation of the a.m. and p.m. suffixes (*am*, *pm*, *a*, *p*, etc.) with or without seconds, and the words *soon* and *midnight*.

Money recognizes any combination of *US*, *CN*, *AUS*, *NZ*, or *HK* with \$, *sterling*, *DM*, *FF*, *USD*, or *CR*; any combination of digits with or without a decimal point.

ID Number recognizes U.S. Social Security Numbers in the format *nnn-nn-nnnn*.

Telephone Number recognizes telephone numbers with or without area codes and includes the symbols () / - and the numerals 0-9; also the words *tel*, *Phone*, *Fax*, *ID*, *No*, and *Num* followed by a colon (:), hyphen (-), slash (/) or period (.).

Text Layout and Quality

The TextBridge control includes a number of options for the recognition of difficult or unusual images and document formats. Documents are frequently printed in columns or in dot matrix text or in low resolution, all of which interfere with accurate character recognition. The TextBridge recognition engine has been trained to accommodate the distortions or difficulties inherent in recognizing these documents. A series of properties in the control, each beginning with the prefix *Opt*, may be independently set to enable each of the recognition features described in the paragraphs below.

Columns or Sections

Many documents which must be read by the TextBridge engine were printed in columns or have sidebars or inserted sections that are not part of the text proper. In all cases, the default behavior of the recognition engine is to read straight across the entire region on each line. If the document is formatted with columns, this behavior will create an unreadable jumble of sentence fragments.

The TextBridge control is capable of automatically detecting columns and other segmented text layouts. After detection, recognition proceeds but follows the flow of text down one column at a time. Detecting and processing columnar text is more time consuming than the regular OCR of single column documents. Therefore, it is advisable to disable this feature except when it is absolutely necessary.

To enable the automatic detection and correction for columns and similar document formatting, set the **OptAutoSegment** property to True before beginning recognition:

```
TextBridge1.OptAutoSegment = True
```

Dot Matrix Printed Text

Some of the most difficult documents to accurately process by OCR are those which were made on a dot matrix printer. Dot matrix text is difficult for OCR engine to recognize because the characters are not continuous, and it is therefore difficult to distinguish between adjacent characters on the printed page. Because it can interfere with recognition of normal text, this option should be disabled except when it is needed.

The TextBridge recognition engine has been specifically trained to be able to more accurately recognize dot matrix print. To enable this feature, set the **OptDotMatrix** property to True before beginning recognition:

```
TextBridge1.OptDotMatrix = True
```

Faxed Documents

Because of the (typically) low resolution, differing vertical and horizontal resolution, and distortion of images that have been faxed, most recognition engines have difficulty accurately recognizing these documents. Using TextBridge's powerful and flexible neural net engine, faxed documents can be much more accurately and reliably recognized.

Two properties may be set to optimize the recognition of faxed documents.

- If the incoming image is known to be a fax document, the **OptFax** property may be set to True. If **OptFax** is True, setting **OptAutoFax** to True will not cause any change in processing as all images will be processed as faxed images.

```
TextBridge1.OptFax = True
```

- If the quality of the incoming document is unknown, the **OptAutoFax** property may be set to True to allow the recognition engine to determine if the image is faxed.

```
TextBridge1.OptAutoFax = True
```

An alternative property is available that combines the features of both the **OptFax** and the **OptAutoSetDegrade** (see below) properties. The **ICRMode** property specifies the type relative readability of an incoming image:

- 0 Standard -- Normal documents
- 1 Distorted -- Faxed or otherwise low resolution or distorted images
- 2 Unknown -- Allows the engine to determine image quality

When recognizing faxed images, enable this feature of TextBridge by setting the **ICRMode** property:

```
TextBridge1.ICRMode = 1
```

Distorted or Degraded Documents

The process of converting old or much-handled documents to electronic form naturally carries over many of the shortcomings of the original documents to the electronic document. For example, the text on a page that has been wrinkled or wetted will be less clear than it was originally and will therefore be more difficult to accurately recognize.

The optical recognition of this degraded text requires additional processing time and greater flexibility in the recognition of individual characters. Therefore, the TextBridge control can be instructed to determine if the document that is being read requires this additional processing effort. To enable the automatic detection of and correction for degraded document text, set **OptAutoSetDegrade** to True:

```
TextBridge1.OptAutoSetDegrade = True
```

Halftone Shading

Halftone shading, as opposed to dot shading, cannot be removed by most image enhancement applications and will dramatically reduce OCR efficiency and accuracy. The TextBridge engine has been trained to identify areas of halftone shading and to take additional processing steps to recognize the text in those regions. This process does require some additional processing time, so should be disabled except when required.

To enable the recognition of halftone shaded documents, set the **OptRemoveHalftone** property to True. To disable this feature, set **OptRemoveHalftone** to False.

```
TextBridge1.OptRemoveHalftone = True
```

Inverted Text

Few recognition engines are capable of processing inverted (white text on black background) images. If other processing is to be applied to these images, ImageBASIC offers options for the correction of inverted documents. Both the ScanFix control and the TMSSequoia Display control can correct inverted pages. The ScanFix control can also find and correct inverted regions.

If you prefer, the TextBridge control can be instructed to expect inverted image data, and the engine will perform the correction before OCR begins. To enable this feature, set the **OptInvert** property to True. When processing standard images (black text on a white background), be sure to disable this feature by setting **OptInvert** to False.

```
TextBridge1.OptInvert = True
```

To maintain backward compatibility, the **thInvert** property has been retained. This is another Boolean property that act identically to the **OptInvert** property. If either one of these properties is True, TextBridge will expect an inverted original.

Rotated Originals

Under many conditions, images are supplied for processing in a variety of orientations -- upside-down or landscape, for example. When processing these images using other ImageBASIC controls, both ScanFix and TMSSequoia Display can rotate the image to the correct orientation. Using the display control, images may be manually corrected, while the ScanFix control can be instructed to automatically detect and correct for rotated images.

- If incoming images are known to be of a certain orientation, recognition will proceed more efficiently if the correction that needs to be applied is specified by setting the **OptRotation** property:

```
TxtBrg1.OptRotation = 1 ' 90 Degrees
```

Options for this enumerated property are as follow:

- 0 No Rotation
 - 1 90 Degrees
 - 2 180 Degrees
 - 3 270 Degrees
- The TextBridge control is also capable of making automatic correction when supplied with rotated documents. If the orientation is unknown, or if the images need to be positioned to portrait, one of the following properties may be set to True. For maximum accuracy, automatic orientation requires at least two lines of text and some white space on the left and right of the text:
 - OptAutoUpright Detects and rotates the image to upright

OptAutoPortrait Rotates the image to portrait

Skewed Originals

Documents are commonly skewed -- i.e., rotated to some degree not included in the primary rotations of 90, 180 and 270 -- during scanning. Typically, the accuracy and efficiency of OCR rapidly degrades above skews of 5%. Therefore, the TextBridge control contains the necessary functionality to deskew an image before it is recognized.

To enable automatic deskewing, set the **OptAutoDeskew** property to True:

```
TextBridge1.OptAutoDeskew = True
```

Verification and Training

As TextBridge is performing recognition on a region or document, the engine can request verification or correction for any word or character that it cannot recognize or that is assigned a confidence below a user-definable level. Using this verification process, TextBridge's neural net technology learns to recognize difficult or uncommon documents and fonts. If desired, this training information can be saved and applied to all future recognition.

Similarly, even when performing recognition without any user intervention, the engine continues to learn on its own. As words are recognized and reported, any slightly uncertain characters within those words are learned and stored. This learning can also be saved for all future recognition attempts, or it can be discarded at the termination of the program.

In the verification process, the TextBridge engine supplies a bitmap of the actual image region which it doubts and requests the input of the correct text. This dialog between the application and the user is controlled through the **Verify** event, as detailed below.

Verification

Verification is the correction or acceptance of the recognition engine's attempt to identify a character or word in an image. It is also the primary method of training for the engine, but the verification information can be stored only temporarily and discarded when the application terminates.

The interaction between the user and the engine takes place in the **Verify** event. This event is triggered during recognition for each character that falls below a user-definable confidence level. Therefore, the following properties must be set

before recognition initiates to specify minimum confidence levels and reporting capabilities:

Verify	If True, the Verify event is enabled. If False, the event will not occur, regardless of the values of the following properties.
ThresholdVerify	Specifies the minimum confidence level below which a character will trigger the Verify event if the Verify property is True.
ThresholdAccept	Specifies the minimum confidence level below which a character will be marked as questionable.
ThresholdReject	Specifies the minimum confidence level below which a character will be marked as noise.

When a recognition attempt is made, the **Verify** event will occur once for each character that falls below the minimum confidence level specified in the **ThresholdVerify** property. This event is designed to provide the engine's best guess at the uncertain character and to accept corrected or verified output from the operator.

Enabling the Verify Event

The **Verify** event must be enabled by setting the **Verify** property to True. If enabled, this event occurs during recognition. The TextBridge control has several other properties that can be set by the user to determine the level of verification the engine requests. The values of these properties are used in comparison to a value from 0 to 999 that is assigned to each character reflecting the level of confidence the engine has in that interpretation.

- The **Verify** event is triggered each time a character in the return string is assigned a confidence below the minimum value specified in the **ThresholdVerify** property.
- During the **Verify** event, the TextBridge engine supplies a small segment of image data. The image that is available is the portion of the recognition region containing the uncertain character.

The image segment may be displayed in an ImageBASIC display control by specifying the TextBridge control in the display control's **ImageDataSource** property.

If the confidence value of a character is above the **ThresholdAccept** property value, it is shown without highlighting in the image.

If the confidence value is between **ThresholdAccept** and **ThresholdReject** the image data will highlight the character.

If the confidence level is below **ThresholdReject** the recognition engine will insert a noise or questionable character.

- The engine supplies its best guess at the correct interpretation of the word containing the uncertain character in the **BestGuess** parameter.
- The operator is allowed to accept the engine's interpretation or to provide a correction in the **Correction** parameter to the event.
- How the recognition engine uses the corrected text, if any, and whether the information is saved is specified by setting the **Action** parameter.
- If the **TrainFileName** property is set to a valid path and file name and the **SaveTraining** property is True, all user input from this event is written to the named file and can be used in future recognition attempts. To enable the use of the learning file in the future, the **TrainFileName** property must specify the file, and the **LoadTraining** property must be True.

Enabling Training

The TextBridge engine will record and maintain all information and instructions entered during the **Verify** event. To enable the perpetuation of this information, the following criteria must be met:

- 1) The **TrainFileName** must be set to a valid path and file name. If the file does not exist, it will be created. If the file does exist it will be appended. The control will not create a nonexistent subdirectory.
- 2) The **SaveTraining** property must be set to True.

Note that the TextBridge engine continues to learn on its own even without user input. For this reason, the training file can become too large or complex when processing many documents. For this reason, the **SaveTraining** property may be set to False to disable the automatic updating of the training file, but the information already contained in the file will be used during recognition if the **LoadTraining** property is True.

Using the Verify Event to Train

When the **Verify** event occurs, the following information is supplied:

- 1) The TextBridge control can be a source of image data for a display control. For example, a TMSSequoia Display control may be used to show the region of the image that contains the uncertain character or word:

```
TMSDisp2.ImageDataSource = TextBridge1.Link
```

In the image that is provided by the TextBridge control, the characters or groups that are uncertain will be visibly different from the other characters.

2) The following parameters to the event are available:

BestGuess	Reports a string of the engine's OCR result for the displayed image
SuspectStart	Reports the starting position of the questionable character in the BestGuess string
SuspectLength	Reports the number of characters in the suspect group
Top	Reports the image pixel position of the top of a rectangle bounding the uncertain group
Bottom	Reports the image pixel position of the bottom of a rectangle bounding the uncertain group
Left	Reports the image pixel position of the left edge of a rectangle bounding the uncertain group
Right	Reports the image pixel position of the right edge of a rectangle bounding the uncertain group
Correction	A string which should be set to the corrected suspect characters as reported in the <i>BestGuess</i> string
Action	<p>Set to one of the following constants to accept the current OCR text or to change the text:</p> <ul style="list-style-type: none">0 Accept Group1 Replace Group2 Join Right3 Join Left4 Accept and Move Right5 Accept and Move Left6 Undo7 Accept Word8 Accept Word No Train <p>See "Action Parameter Options" on page 30 for details on each of these options.</p>
Cancel	If set to True during this event, verification will be canceled without performing any update.

Action Parameter Options

The term *group* refers to any collection of black pixels which is contiguous. Normally, this refers to any single character. Sometimes, however, several characters are interpreted as a single group, and on other occasions one letter is broken into more than one part. In instances such as these, the options of joining

adjacent groups and allowing the user to type in the correct interpretation of the entire string of characters are allowed.

0--Accept Group

The character(s) in the *BestGuess* parameter are accepted as they appear. Any entry made in the *Correction* parameter will be ignored.

Recognition will now continue, and the **Verify** event will occur again for the next character which falls below **ThresholdVerify**

1--Replace Group

This action causes the engine to replace the group that triggered this event with the string to which the *Correction* parameter is set.

This correction will be saved for future recognition events if a valid file is specified in the **TrainFileName** property and **SaveTraining** is True.

2--Join Right

Causes the current group to be joined to the group immediately to its right. This feature will be used when a single letter is split in the image and is considered to the two characters by the recognition engine. The **Verify** event will occur again for the newly joined group. No verification information or training is created by this action.

3--Join Left

Causes the current group to be joined to the group immediately to its left. This feature will be used when a single letter is split in the image and is considered to the two characters by the recognition engine. The **Verify** event will occur again for the newly joined group. No verification information or training is created by this action.

4--Accept and Move Right

This action causes the string to be accepted as originally presented in the *BestGuess* parameter and triggers the **Verify** event again for the group to the right of the current group.

5--Accept and Move Left

This action causes the string to be accepted as originally presented in the *BestGuess* parameter and triggers the **Verify** event again for the group to the left of the current group.

6--Undo

Causes the most recently entered information to be undone. For example, if in the previous **Verify** event, the *Action* parameter was set to *1--Replace Group* and the training file was updated based on that entry, selecting the **Undo** action will remove that entry from the training file and set the output string back to the engine's original interpretation.

7--Accept Word

This action causes the string to be accepted as originally presented in the *BestGuess* parameter. This confirmation will be saved for future recognition events if a valid file is specified in the **TrainFileName** property and **SaveTraining** is True.

8--Accept Word No Train

This action causes the string to be accepted as originally presented in the *BestGuess* parameter. This confirmation will not be saved for future recognition events even if a valid file is specified in the **TrainFileName** property and **SaveTraining** is True.

Chapter 3 : Reference

Reference to Properties, Methods and Events

AboutBox Method

Definition:	Displays a message box showing version and copyright information when queried.
Syntax:	<code>TxtBrgl .AboutBox</code>
Design Access:	Read-only
Runtime Access:	Read-only
Comments:	The message box is application modal and contains a single OK button. Clicking the button will close the message box.

Active Property

Definition:	<p>If set to True at design time, the control will fully initialize and verify licensing immediately upon initialization of the application at application.</p> <p>If set to False at design time, full initialization of the control will be delayed at initialization of the application at runtime. In this case, this property must be explicitly set to True at runtime before the control is used.</p>
Data Type:	Boolean
Design Access:	Read/Write
Runtime Access:	Read/Write (see limits below)
See Also:	"Licensing Configuration and Verification on page 2
Comments:	<p>If this property is set to True (the default) at design time, the control is fully initialized and licensing is verified immediately upon initialization of the application at runtime. The related technology libraries are loaded and the control is ready to be used.</p> <p>If this property is set to False at design time, the control will only partially initialize when the application loads at runtime. By delaying these two actions, the application should be able to load more quickly:</p> <ol style="list-style-type: none">1) The related technology libraries for the control will not be loaded.2) The licensing server will not verify an available token for the control.

If the control initializes with **Active** set to False, this property must be explicitly set to True by the application. Until **Active** is set to True, the control will ignore all instructions to it.

If the control fails to find a license token, the **Active** property will be automatically set to False. The application can check this value on Form Load to determine if each control is licensed and can be used.

Note: The **Active** property cannot be set to False by the application at runtime. The only time that the **Active** property can be changed from True to False at runtime is when a licensing token is not found. Once locked, all license tokens will remain locked until the application terminates.

AddRegion Method

Definition:	Allocates a new region for multiple region processing.
Parameters:	None
Syntax:	<code>nRegIndex = TxtBrg1.AddRegion</code>
Data Type:	Long Integer
Return Values:	Index value of the newly allocated region
See Also:	DeleteRegion Method, RegCount Property
Comments:	When this method is successfully executed, the RegCount property is incremented by one. A new element of the region definition properties, listed below, is allocated and all coordinates for the new region are set to 0 (zero). The new region may be modified by selecting the new region through the RegIndex property and changing the value of the region definition properties: RegLeft RegTop RegRight RegBottom

Completion Event

Definition:	Occurs frequently during recognition and provides percentage completed and the option to cancel recognition.				
Parameters:	<table><tr><td>Percent</td><td>Percentage completion of current OCR attempt</td></tr><tr><td>Cancel</td><td>If set to True, this OCR attempt is canceled</td></tr></table>	Percent	Percentage completion of current OCR attempt	Cancel	If set to True, this OCR attempt is canceled
Percent	Percentage completion of current OCR attempt				
Cancel	If set to True, this OCR attempt is canceled				
Comments:	<p>The only way to cancel a recognition attempt is to set the Cancel parameter to this event to True during the event. The cancellation may be passed to the event as a global variable set in a button click or through some other method, as shown here:</p> <pre>Private Sub cmdCancel_Click() gnStop = True End Sub Private Sub TxtBrgl_Completion(ByVal Percent As Integer, Cancel As Boolean) DoEvents Cancel = gnStop End Sub</pre>				

DebugPath Property

Definition:	If set to a valid file name, the control will generate debug information and write it to the file.
Data Type:	String
Design Access:	Read/Write
Runtime Access:	Read/Write
See Also:	Progress Event
Comments:	The debug information that is supplied here is related mainly to the OCR process. This information may be requested by a technical support representative. The information written to this file is detailed and reflects the low-level processes of the engine.

DeleteRegion Method

Definition:	Deletes the current region element as specified in the RegIndex property.
Parameters:	None
Syntax:	<code>TxtBrgl.DeleteRegion</code>
Data Type:	Long Integer
Return Values:	None

See Also: AddRegion Method, RegIndex Property

Comments: When this method is successfully executed, the **RegCount** property is updated, and the **RegIndex** property is changed only if necessary.

Error Event

Definition: Occurs for each error internal to the control.

Parameters:

Number	A long error code that identifies the error
Description	Descriptive string of the error
SCode	A composite long number indicating the severity of the error, the facility code, the origin of the error, and the native error code
Source	Descriptive string of the source of the error
HelpFile	Suggested help file name that should have a detailed explanation of the error
HelpContext	Context ID of the appropriate topic in the help file named above
CancelDisplay	If set to True during this event, the standard error dialog will not be displayed

ICRMode Property

Definition: Specifies whether the image is standard or was generated by a fax server or machine. Faxed documents require additional processing to compensate for the typically lower resolution and image quality.

Data Type: Enumerated

Design Access: Read/Write

Runtime Access: Read/Write

Possible Values:

0	Standard
1	Degraded
2	Unknown

See Also: OptAutoSetDegrade Property, OptAutoFaxProperty

Description: This property is superseded by the **OptAutoSetDegrade** and **OptAutoFax** properties but has been retained for backward compatibility.

0--Standard is the preferred setting for normal documents.

1--Degraded enhances recognition of faxed or other low resolution documents but can slow processing and reduce accuracy if used when unnecessary.

2--*Unknown* may be specified if the quality of the incoming documents is unknown. The recognition engine will determine the optimum setting for each image.

ImageDataSource Property

- Definition:** Specifies the ImageBASIC control that will supply image data for OCR by this control. Valid only when the **InputFrom** property is set to 0--*ImageDataSource*
- Data Type:** String
- Syntax:** TextBridge1.ImageDataSource = TMSDisp1
- Design Access:** Read/Write
- Runtime Access:** Read/Write
- See Also:** InputFrom Property, RegionSource Property
- Comments:** **ImageDataSource** must specify an ImageBASIC control that can supply image data if OCR is to be attempted and the **InputFrom** property is 0--*ImageDataSource*
- When an ImageBASIC control is added to a Form at design time, the **ImageDataSource** property is automatically populated with a source ImageBASIC control if one already exists on the Form.
- At runtime, this property may be set to the **Link** value of any ImageBASIC control that is an image source; for example,
- ```
TextBrg1.ImageDataSource = TMSDisp1.Link
```

### ***InputFileName Property***

- Definition:** Specifies the path and file name of the image file to OCR. Valid only when the **InputFrom** property is set to 1--*File*.
- Data Type:** String
- Design Access:** Read/Write
- Runtime Access:** Read/Write
- See Also:** InputFrom Property, OutputFileName Property
- Comments:** If a multiple-page file is selected, all pages of the file will be processed. The TextBridge control can read and process the following file types:
- TIFF Group 4
  - TIFF Group 3
  - TIFF Group 3 Modified
  - TIFF Uncompressed
  - TIFF Packed

## ***ImageResolution Property***

|                        |                                                                                                                                                     |
|------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Definition:</b>     | Reports the DPI resolution of the image currently being processed by the control.                                                                   |
| <b>Data Type:</b>      | Long                                                                                                                                                |
| <b>Design Access:</b>  | Not Available                                                                                                                                       |
| <b>Runtime Access:</b> | Read-only                                                                                                                                           |
| <b>See Also:</b>       | InputFrom Property, ImageDataSource Property                                                                                                        |
| <b>Comments:</b>       | This property will report a valid value only when an image is available to the control. This property is informational only and cannot be modified. |

## ***InputFrom Property***

|                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Definition:</b>      | Specifies the source of image data for recognition -- from another ImageBASIC control or directly from file.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| <b>Data Type:</b>       | Enumerated                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| <b>Design Access:</b>   | Read/Write                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| <b>Runtime Access:</b>  | Read/Write                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| <b>Possible Values:</b> | 0 ImageDataSource<br>1 File                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <b>See Also:</b>        | ImageDataSource Property, InputFileName Property, RegionSource Property, OutputTo Property                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| <b>Comments:</b>        | <p>If set to <i>0--ImageDataSource</i>, an ImageBASIC control that can serve as an image source must be specified in the <b>ImageDataSource</b> property. The ImageBASIC control that most commonly serves as an image source is the TMSSequoia Display control.</p> <p>If set to <i>1--File</i>, the <b>InputFileName</b> must specify a valid path and file name.</p> <p>All recognition attempts may be performed on either the entire input page, using the <b>OCRPage</b> method, or on a defined region of the image, using the <b>OCRRegion</b> method.</p> <p>When the <b>OCRRegion</b> method is used, either the <b>RegionSource</b> property can specify an ImageBASIC control that can supply region coordinates, or the TextBridge control's <b>RegLeft</b>, <b>RegTop</b>, <b>RegRight</b> and <b>RegBottom</b> properties can specify the region. The ImageBASIC Display controls are the primary source of region coordinates, which correspond to the current Working Region of the Display.</p> |

### ***Invert Property***

|                        |                                                                                                                                                                         |
|------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Definition:</b>     | If True, the control will invert the received image data before any other processing.                                                                                   |
| <b>Data Type:</b>      | Boolean                                                                                                                                                                 |
| <b>Design Access:</b>  | Read/Write                                                                                                                                                              |
| <b>Runtime Access:</b> | Read/Write                                                                                                                                                              |
| <b>See Also:</b>       | OptInvert Property (Offers identical functionality)                                                                                                                     |
| <b>Description:</b>    | This must be set to True if the incoming image data is already inverted (white text on a black background), because the recognition engine cannot process inverse text. |

### ***Language Property***

|                        |                                                                                                                                                                                          |
|------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Definition:</b>     | Specifies which of the installed languages lexicons will be used during recognition.                                                                                                     |
| <b>Data Type:</b>      | Enumerated                                                                                                                                                                               |
| <b>Design Access:</b>  | Read/Write                                                                                                                                                                               |
| <b>Runtime Access:</b> | Read/Write                                                                                                                                                                               |
| <b>See Also:</b>       | LexicalClass Property, LexicalMode Property                                                                                                                                              |
| <b>Comments:</b>       | This property should be set to use different libraries to accord with the language in which the image text being recognized was written. Valid options for this property are as follows: |

- 0 English
- 1 German
- 2 French
- 3 Spanish
- 4 Italian
- 5 Dutch
- 6 Norwegian
- 7 Finnish
- 8 Danish
- 9 Portuguese
- 10 Reserved
- 11 Russian
- 12 English & German

Although the TextBridge control is shipped with all of the currently available languages, one or more of these files may have been removed from your system. Therefore, be careful to set this property to an installed language file, or the recognition engine will not function properly.

## LexicalClass Property

**Definition:** Specifies one of the built-in field types to optimize recognition.

**Data Type:** Enumerated

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**See Also:** LexicalMode Property, Language Property, LexiconFileName Property

**Comments:** At design time, this property is set to any of its valid values through a dialog that is displayed when the property is clicked in the Properties Window. Each option may be marked with a check to select any number of field classes.

At runtime, the property is set through the constants listed below with the valid options. To select no lexical class, set this property to 0 (zero). More than one of these classes may be selected by adding together the constants, as in this example:

```
nLexSum = IBBF_LEXCLS_NUMBER + IBBF_LEXCLS_TIME
```

```
TxtBrgl.LexicalClass = nLexSum
```

The options for this property and the constant for each option are as follows:

|                  |                           |
|------------------|---------------------------|
| Main             | IBBF_LEXCLS_MAINLEX       |
| Number           | IBBF_LEXCLS_NUMBER        |
| Roman Numeral    | IBBF_LEXCLS_ROMAN_NUMERAL |
| Date             | IBBF_LEXCLS_DATE          |
| Time             | IBBF_LEXCLS_TIME          |
| Money            | IBBF_LEXCLS_MONEY         |
| ID Number        | IBBF_LEXCLS_ID_NUMBER     |
| Telephone Number | IBBF_LEXCLS_TELEPHONE     |
| Terminal Dot     | IBBF_LEXCLS_TERMINAL_DOT  |

*Main* causes recognition to proceed using only the language lexicon but not a user-defined lexicon.

*Number* recognizes the numerals 0-9, fractions, mathematical symbols and numbers ending *int*, *nd*, *rd*, or *th*; for example, 1st, 2nd, etc.

*Roman Numeral* recognizes all characters that are valid in Roman numerals except single characters, three letter combinations beginning with *M* or *m* (to avoid erroneous recognition of words like *nix*), and the combination *LI* or *li* to avoid confusion with *ll*, *ll*, or *ll*.



*Date* recognizes any combination of a date, month, and year in numeral or with the month spelled or abbreviated. Both U.S. and European formats are recognized.

*Time* recognizes 12-hour and 24-hour times including any variation of the a.m. and p.m. suffixes (*am*, *pm*, *a*, *p*, etc.) with or without seconds, and the words *noon* and *midnight*.

*Money* recognizes any combination of *US*, *CN*, *AUS*, *NZ*, or *HK* with \$, sterling, *DM*, *FF*, *USD*, or *CR*; any combination of digits with or without a decimal point.

*ID Number* recognizes U.S. Social Security Numbers in the format *nnn-nn-nnnn*.

*Telephone Number* recognizes telephone numbers with or without area codes and includes the symbols ( ) / - and the numerals 0-9; also the words *Tel*, *Phone*, *Fax*, *ID*, *No*, and *Num* followed by a colon (:), hyphen (-), slash (/) or period (.)

## ***LexicalMode Property***

**Definition:** Specifies the level to which the recognition engine will attempt to match the data in the lexicon(s) being used.

**Data Type:** Enumerated

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**Possible Values:**

|   |                       |
|---|-----------------------|
| 0 | No Lexical Processing |
| 1 | Preference            |
| 2 | Absolute              |

**See Also:** *LexiconFileName* Property, *LexicalClass* Property, *Language* and *Field Type Specification* on page 15

**Comments:** *0--No Lexical Processing* instructs the engine to ignore the information in both the standard language lexicon and the user-defined lexicon, if one has been specified.

*1--Preference* instructs the engine to accept words returned during OCR if they are similar to an entry in either the standard language lexicon or the user-defined lexicon. Verification will still be requested for characters falling below the

**ThresholdAccept** level. A word that is found in a lexicon will be assigned a higher confidence, so verification is less likely.

*2--Absolute* instructs the recognition engine to request verification for any word that it does not find in either the standard lexicon, specified with the **Language** property, or the user-defined lexicon, specified in the **LexiconFileName** property. If a word is accepted or a correction made that does

not match an entry in one of the available dictionaries, that word will be reported as unrecognizable.

### ***LexiconFileName Property***

- Definition:** Specifies fully qualified path and file name of the user-defined lexicon that will be used during OCR.
- Data Type:** String
- Design Access:** Read/Write
- Runtime Access:** Read/Write
- See Also:** LexicalMode Property, LexicalClass Property, Language and Field Type Specification on page 15
- Comments:** A user-defined lexicon is simply a plain text file with one entry per line. The exact use to which the lexicon will be put is specified by the **LexicalMode** and **LexicalClass** properties. Refer to "User-Defined Lexicon" on page 16 for details.

### ***Link Property***

- Definition:** Reports the Link ID calculated for this control at its creation.
- Data Type:** String
- Syntax:** `TMSDispl.ImageDataSource = Txtbrdg1.Link`
- Design Access:** Not Available
- Runtime Access:** Read-only
- Comments:** Each ImageBASIC control is assigned a unique Link ID at its creation. This Link ID can be specified in the **ImageDataSource**, **DisplaySource**, **RegionSource** and **AnnoteSource** properties of various ImageBASIC controls. These source properties specify the ImageBASIC control that is supplying information or services to a control. The TextBridge control will provide image data only during the **Verify** event. The image that is available at this time is the portion of the original image that contains the suspect character(s). Refer to section "Verification" on page 26 for details on this process.

### ***LoadTraining Property***

- Definition:** If True, any learning file specified in the TrainFileName property will be loaded into memory.
- Data Type:** Boolean
- Design Access:** Read/Write
- Runtime Access:** Read/Write

**See Also:** TrainFileName Property, VerifyEvent, LexicalMode Property

**Comments:** If a **TrainFileName** is specified and populated through the verification process, the TextBridge control will read from that file to improve recognition accuracy. The learning file will be loaded only if the **LoadTraining** property is True and a valid file is specified.

### **OCRPage Method**

**Definition:** Initiates a recognition attempt on the current image page.

**Parameters:** None

**Syntax:** TextBridge1.OCRPage

**Data Type:** String

**Return Values:** OCR result string

**See Also:** OCRRegion Method, InputFrom Property, OutputTo Property

**Comments:** The image on which recognition will be performed is specified as follows:

If the **InputFrom** property is set to *0--ImageDataSource*, the image for OCR will be read from the ImageBASIC control specified in the **ImageDataSource** property.

If the **InputFrom** property is set to *1--File*, the image for OCR will be read from the file(s) specified in the **InputFileName** property.

The entire input page will be processed even if the **RegionSource** property specifies a valid source of region coordinates.

### **OCRRegion Method**

**Definition:** Initiates a recognition attempt on the current image region using the current defined region coordinates.

**Parameters:** None

**Syntax:** TextBridge1.OCRRegion

**Data Type:** String

**Return Values:** OCR result string

**See Also:** OCRPage Method, RegionSource Property, InputFrom Property, OutputTo Property

**Comments:** The image on which recognition will be performed is specified as follows:

If the **InputFrom** property is set to *0--ImageDataSource*, the image for OCR will be read from the ImageBASIC control specified in the **ImageDataSource** property.

If the **InputFrom** property is set to *File*, the image for OCR will be read from the file(s) specified in the **InputFileName** property.

The region of this image to be processed can be specified in either of two ways:

- 1) The **RegionSource** property may be set to specify any ImageBASIC control that can supply region coordinates. As of this writing, the Display controls are the only controls that can define a Working Region.

By default, if the image data is read from a Display control, the Working Region defined in that control will be the region that is read by OCR.

- 2) The TextBridge control can define the region using the following properties. These properties are applied only if **RegionSource** is none:

RegBottom

RegLeft

RegRight

RegTop

If multiple regions are defined (see **RegCount** property), all regions will be recognized.

### ***OptAutoDeskew Property***

- Definition:** If True, the recognition engine will attempt to identify and correct skewed text before recognition begins. If False, the text will be read as presented.
- Data Type:** Boolean
- Design Access:** Read/Write
- Runtime Access:** Read/Write
- See Also:** OptAutoUpright Property
- Description:** The recognition of text is degraded at skews of more than a few degrees. Therefore, deskewing should be enabled unless the incoming images are known to be straight.

### ***OptAutoFax Property***

- Definition:** If True, the recognition engine will determine if the incoming image is fax quality and will automatically adjust to compensate.
- Data Type:** Boolean
- Design Access:** Read/Write

**Runtime Access:** Read/Write

**See Also:** OptFax Property, OptAutoSetDegrade Property

**Description:** Faxed images are commonly slightly distorted and are frequently transmitted with different vertical and horizontal resolutions. When performing OCR on these images, TextBridge can compensate for the difficulty introduced by these characteristics if the **OptAutoFax** or **OptFax** property is True.

The difference between the two options is that if **OptFax** is True, all incoming images are processed as if they were faxed, while if **OptAutoFax** is True, the engine will determine if the images were faxed. **Note:** Only one of these properties should be set to True at any given time.

### ***OptAutoPortrait Property***

**Definition:** If True, incoming images will be analyzed to determine if the page or region is portrait (i.e., the sides are longer than the width) and the images will be rotated to portrait if necessary.

**Data Type:** Boolean

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**See Also:** OptAutoUpright Property

**Description:** Most documents are printed in portrait mode, but many scanning processes feed the documents in landscape mode to speed the paper feed. The text in the image that is being read must be upright, and TextBridge will rotate the incoming image to portrait before further processing if the **OptAutoPortrait** property is True.

**Note:** The image may be up-side-down even if it is portrait. Therefore, the **OptAutoUpright** property may be used instead of **OptAutoPortrait** if necessary.

### ***OptAutoSegment Property***

**Definition:** If True, the recognition engine will attempt to identify columns and segmented text and will output the OCR results accordingly.

**Data Type:** Boolean

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**See Also:** OptAutoUpright Property

**Description:** The normal procedure of most OCR engines is to read completely across the page on each line during recognition. If the text is in columns, this will result in essentially unreadable sentence fragments. If **OptAutoSegment** is True, TextBridge will attempt to identify columns and text sections such as sidebars or captions and output the recognition results as the text actually flows.

### ***OptAutoSetDegrade Property***

**Definition:** If True, TextBridge will automatically determine if the image text is degraded and will adjust as necessary.

**Data Type:** Boolean

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**See Also:** ICRMode Property

**Description:** Setting this property to True is essentially identical to setting the **ICRMode** property to 2--*Unknown*. If the text is known to be either degraded or not, preprocessing will be quicker if **ICRMode** is set to the appropriate value.

### ***OptAutoUpright Property***

**Definition:** If True, TextBridge will attempt to identify the upright position of the text on the input image and rotate the image to correct for inappropriate rotation before recognition is started.

**Data Type:** Boolean

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**See Also:** OptAutoPortrait Property, OptAutoDeskew Property

**Description:** Determination of the actual upright position of text is based largely on the relative abundance of ascenders relative to descenders on the letters. For this reason, a relatively large segment of text -- at least two lines of text with white space on the left and right -- must be available in the region for the engine to correctly determine what is upright.

**Note:** The automatic determination of the upright position of the image text requires some additional processing time, and, therefore, this option should be disabled unless it is necessary.

### ***OptDoRight Property***

|                        |                                                                                                                                                                     |
|------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Definition:</b>     | If True, TextBridge is configured to automatically detect and compensate for a series of image defects.                                                             |
| <b>Data Type:</b>      | Boolean                                                                                                                                                             |
| <b>Design Access:</b>  | Read/Write                                                                                                                                                          |
| <b>Runtime Access:</b> | Read/Write                                                                                                                                                          |
| <b>See Also:</b>       | See Below                                                                                                                                                           |
| <b>Description:</b>    | Enabling this feature is equivalent to setting the following properties to True:<br>OptAutoDeskew<br>OptAutoUpright<br>OptAutoFax<br>OptAutoSegment<br>OptDotDetect |

### ***OptDotDetect Property***

|                        |                                                                                                                                                                                                                                      |
|------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Definition:</b>     | If True, TextBridge will automatically determine if the incoming image is dot matrix print and will adjust as necessary.                                                                                                             |
| <b>Data Type:</b>      | Boolean                                                                                                                                                                                                                              |
| <b>Design Access:</b>  | Read/Write                                                                                                                                                                                                                           |
| <b>Runtime Access:</b> | Read/Write                                                                                                                                                                                                                           |
| <b>See Also:</b>       | OptDotMatrix Property                                                                                                                                                                                                                |
| <b>Description:</b>    | If the incoming image is known to be dot matrix or known not to be dot matrix, preprocessing will proceed optimally if <b>OptDotDetect</b> is left at the default of False and <b>OptDotMatrix</b> is set to specify the image type. |

### ***OptDotMatrix Property***

|                        |                                                                                                                                                                                      |
|------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Definition:</b>     | If True, TextBridge will expect incoming images to contain dot matrix text. If False, normal recognition will proceed, but dot matrix text will be poorly recognized.                |
| <b>Data Type:</b>      | Boolean                                                                                                                                                                              |
| <b>Design Access:</b>  | Read/Write                                                                                                                                                                           |
| <b>Runtime Access:</b> | Read/Write                                                                                                                                                                           |
| <b>See Also:</b>       | OptDotDetect Property, OptAutoSetDegrade Property                                                                                                                                    |
| <b>Description:</b>    | Because of its nature, dot matrix text requires special processing parameters to maximize recognition accuracy. TextBridge includes additional processing steps when recognizing dot |

matrix, but this property must be set to True to enable this feature.

**Note:** Because of the additional processing performed when this property is True, recognition will proceed slightly slower than usual. In addition, regular text may not be recognized as accurately. Therefore, disable dot matrix recognition unless it is required.

### ***OptFax Property***

**Definition:** If True, the recognition engine will adjust to compensate for the particular difficulties of faxed images.

**Data Type:** Boolean

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**See Also:** OptAutoFax Property, OptAutoSetDegrade Property

**Description:** Faxed images are commonly slightly distorted and are frequently transmitted with different vertical and horizontal resolutions. When performing OCR on these images, TextBridge can compensate for the difficulty introduced by these characteristics if the **OptAutoFax** or **OptFax** property is True.

The difference between the two options is that **OptFax** is True, all incoming images are processed as if they were faxed, while if **OptAutoFax** is True, the engine will determine if the images were faxed. Only one of these properties should be set to True at any given time.

### ***OptFullOrientation Property***

**Definition:** If True, incoming images will be analyzed for to determine if they are both portrait and upright. The image will be rotated as necessary to make the text on the image upright.

**Data Type:** Boolean

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**See Also:** OptAutoUpright Property, OptAutoPortrait Property

**Description:** Setting this property to True is equivalent to setting the **OptAutoUpright** and **OptAutoPortrait** properties to True.

### ***OptInvert Property***

**Definition:** If True, the control will invert the received image data before any other processing.



**Data Type:** Boolean  
**Design Access:** Read/Write  
**Runtime Access:** Read/Write  
**Description:** This must be set to True if the incoming image data is already inverted (white text on a black background), because the recognition engine cannot process inverse text.

### ***OptNewspaper Property***

**Definition:** If True, recognition will be optimized for segmentation of scanned newspaper. If False, normal print is expected.  
**Data Type:** Boolean  
**Design Access:** Read/Write  
**Runtime Access:** Read/Write  
**See Also:** OptDotMatrix Property, OptInvert Property, OptAutoSetDegrade Property  
**Description:** This option primarily optimizes for the segmentation (column detection) and text quality of newspaper.

### ***OptRemoveHalftone Property***

**Definition:** If True, TextBridge will remove any halftone shading in the recognition region before OCR begins. If False and halftone shading is present, OCR accuracy will be significantly degraded.  
**Data Type:** Boolean  
**Design Access:** Read/Write  
**Runtime Access:** Read/Write  
**See Also:** OptDotDetect Property, OptNewspaper Property  
**Description:** Halftone shading, in contrast to dot shading, cannot be removed by many image enhancement processes, but it will interfere with accurate OCR. TextBridge can remove or disregard halftone shading when the **OptRemoveHalftone** property is True.  
When halftone removal is enabled, pre-recognition processing requires additional time. It is therefore advisable to disable this feature except when it is required.

### ***OptRotation Property***

**Definition:** Specifies the degree to which the incoming image should be rotated before further processing.  
**Data Type:** Enumerated  
**Design Access:** Read/Write

**Runtime Access:** Read/Write

**Possible Values:** 0 None  
1 90 Degrees  
2 180 Degrees  
3 270 Degrees

**See Also:** OptAutoDeskewProperty, OptAutoUpright Property

**Description:** Options are available in the TextBridge control to rotate the incoming image to any of the primary positions or to allow the control to automatically determine the proper orientation. If the rotation of the incoming images is known, pre-recognition processing is faster if the exact degree of correctional rotation is performed. If the orientation of the incoming images is unknown, the control can perform automatic orientation, enabled through the **OptAutoUpright** property. Rotation of the image is rotation to the primary positions. Small degrees of rotation, or skew, can also be detected and corrected. Skew correction is enabled through the **OptAutoDeskew** property.

### ***OptXDocCharConf Property***

**Definition:** Specifies whether or not character confidence values will be reported in the XDoc output format.

**Data Type:** Boolean

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**Possible Values:** True  
False

**See Also:** OutputFormat Property

**Description:** Applies only if **OutputFormat** is 2--*Enhanced XDoc*. If True, character confidence values will be included. If False, character confidence values will be excluded.

### ***OptXDocLexicalClass Property***

**Definition:** Specifies whether or not lexical class information will be included in the XDoc output.

**Data Type:** Boolean

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**Possible Values:** True  
False

**See Also:** OutputFormat Property

**Description:** Applies only if **OutputFormat** is 2--*Enhanced XDoc*. If True, lexical class information will be included. If False, lexical class information will be excluded.

### ***OptXDocWordBox Property***

**Definition:** Specifies whether or not the bounding coordinates for each word will be included in the XDoc output.

**Data Type:** Boolean

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**Possible Values:** True  
False

**See Also:** OutputFormat Property

**Description:** Applies only if **OutputFormat** is 2--*Enhanced XDoc*. If True, word coordinates will be included. If False, word coordinates will be excluded.

### ***OptXDocWordConf Property***

**Definition:** Specifies whether or not word confidence values will be included in the XDoc output.

**Data Type:** Boolean

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**Possible Values:** True  
False

**See Also:** OutputFormat Property

**Description:** Applies only if **OutputFormat** is 2--*Enhanced XDoc*. If True, word confidence values will be included. If False, word confidence values will be excluded.

### ***Orientation Property***

**Definition:** Specifies the degree to which the incoming image should be rotated before further processing.

**Note:** Retained for backward compatibility; now superseded by the **OptRotation** property.

**Data Type:** Enumerated

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**Possible Values:** 0 None  
1 90 Degrees  
2 180 Degrees  
3 270 Degrees

**See Also:** OptRotation Property

**Description:** Options are available in the TextBridge control to rotate the incoming image to any of the primary positions or to allow the control to automatically determine the proper orientation. If the rotation of the incoming images is known, pre-recognition processing is faster if the exact degree of correctional rotation is performed. If the orientation of the incoming images is unknown, the control can perform automatic orientation, enabled through the **OptAutoUpright** property. Rotation of the image is rotation to the primary positions. Small degrees of rotation, or skew, can also be detected and corrected. Skew correction is enabled through the **OptAutoDeskew** property.

### ***OutputFileAppend Property***

**Definition:** If True, the output file will be appended if it exists. If False, the output file will be overwritten if it exists.

**Data Type:** Boolean

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**See Also:** OutputToProperty, OutputFileNameProperty

**Comments:** This property is valid only if the TextBridge control is writing OCR results directly to file. The **OutputTo** property specifies the reporting of results to either a string property or to a file. When writing to file, the **OutputFileName** property specifies the file which the control will write. Only plain text files, not word processor or other specialized formats, can be appended.

### ***OutputFileName Property***

**Definition:** Specifies the path and file name of the file to write with the recognition results.

**Data Type:** String

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**See Also:** OutputToProperty, OutputFileAppendProperty

**Comments:** This property is valid only if the **OutputTop** property specifies *!--File*. The **OutputFileAppend** property controls whether the text is appended to the file or replaces any existing contents.

## ***OutputFormat Property***

**Definition:** Specifies the format of the recognition output string.

**Data Type:** Enumerated

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**See Also:** OutputFileNameProperty, OutputToProperty

**Comments:** The valid enumerated values for this property are shown in the list below.

**Note:** All output formats numbered six and higher must be written to file instead of to a Visual Basic string variable or property. This is because all of these formats are binary output rather than text, and Visual Basic string properties and variables cannot accept this data.

- 1 None (0 byte file will be created)
- 2 Enhanced XDOC
- 3 Formatted ASCII Text
- 4 XDOC without Format Analysis
- 5 Formatted ISO
- 6 Forms ISO
- 7 Ami Pro 2.0
- 8 Ami Pro 3.0
- 12 ASCII Smart
- 13 ASCII Standard
- 14 ASCII Standard DOS
- 15 ASCII Stripped
- 16 dBase IV 1.0
- 17 DCA/RTF
- 18 Display Write
- 20 Excel for Macintosh
- 21 Excel 3.0
- 22 Excel 4.0
- 32 FrameMaker
- 33 Interleaf
- 34 Lotus 1-2-3
- 38 MultiMate Advantage II
- 43 Postscript
- 44 Professional Write 2.0
- 45 Professional Write 2.2

- 47 RTF (Rich Text Format)
- 48 RTF Macintosh
- 49 RTF MS Word 6.0
- 50 Samna Word IV
- 51 Windows Write
- 53 Word For Windows 2.x
- 54 Word For Windows 6.x
- 55 WordPerfect (Windows)
- 56 WordPerfect 4.2
- 59 WordPerfect 6.0
- 60 WordPerfect 6.1 (Windows)
- 64 WordStar 1.x (Windows)
- 68 HTML (2.0)
- 69 HTML (Netscape)
- 70 HTML (SoftQuad Editor)

XDOC is a marked-up text format. It includes additional information along with the text such as confidence information and character position. For complete details on this output format, please contact our Technical Support staff. A fifty-page document describing XDOC in detail is also available on the BBS.

## ***OutputTo Property***

- Definition:** Specifies the destination for OCR results.
- Data Type:** Integer (Enumerated)
- Design Access:** Read/Write
- Runtime Access:** Read/Write
- See Also:** InputFromProperty, OutputFileNameProperty
- Comments:** Valid options for this property are as follows:

- 0 String
- 1 File

*0--String* causes the recognition results to be reported in the **Result** property.

*1--File* causes the recognition results to be written to the file specified in the **OutputFileName** property.

## ***PageIndex Property***

- Definition:** When reading a multiple-page file directly from disk (i.e., the **InputFrom** property is set to *1--File*), specifies which page to OCR.  
If set to 0 (zero), all pages in the file are read.
- Data Type:** Integer
- Design Access:** Not Available
- Runtime Access:** Read/Write
- See Also:** InputFromProperty, InputFileName Property
- Comments:** If all pages of a multiple-page file are processed, the entire OCR results will be written to a single output file. If each individual page is processed, only plain text files can be appended once the output file exists. Therefore, when processing a multiple-page image file and writing the OCR results to a single word processor or similar format file, the **PageIndex** property should be set to OCR all pages of the image file in a single attempt when the **OCRPage** method is executed.

## ***RecognizeCount Property***

- Definition:** Reports the number of words recognized in the most recent OCR attempt.
- Data Type:** Integer
- Design Access:** Not Available
- Runtime Access:** Read-only
- See Also:** OCRRegion Method, OCRPage Method, Result Property
- Comments:** This property is valid only immediately after a successful recognition attempt. The word count is available after any successful OCR attempt regardless of the output format or destination.

## ***RegBottom Property***

- Definition:** Reports or sets the image pixel height of the region for recognition.
- Data Type:** Long
- Design Access:** Read/Write
- Runtime Access:** Read/Write
- See Also:** OCRRegion Method, RegIndex Property
- Comments:** When the **OCRRegion** method is executed to begin recognition and no **RegionSource** is specified, the TextBridge control's own region definition properties are applied. The **RegCount**



property reports the number of defined regions and the following additional properties are used to define those regions:

|                  |                                                                                                       |
|------------------|-------------------------------------------------------------------------------------------------------|
| <b>RegIndex</b>  | May be set to any integer from 1 to <b>RegCount</b> . When set, the following properties are updated. |
| <b>RegLeft</b>   | Image pixel coordinate of left edge of region                                                         |
| <b>RegTop</b>    | Image pixel coordinate of top edge of region                                                          |
| <b>RegRight</b>  | Image pixel coordinate of right edge of region                                                        |
| <b>RegBottom</b> | Image pixel coordinate of bottom edge of region                                                       |

### ***RegCount Property***

|                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                 |                                                                                                       |                |                                               |               |                                              |                 |                                                |                  |                                                 |
|------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|-------------------------------------------------------------------------------------------------------|----------------|-----------------------------------------------|---------------|----------------------------------------------|-----------------|------------------------------------------------|------------------|-------------------------------------------------|
| <b>Definition:</b>     | Reports the total number of regions defined for multi-region processing.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                 |                                                                                                       |                |                                               |               |                                              |                 |                                                |                  |                                                 |
| <b>Data Type:</b>      | Long                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                 |                                                                                                       |                |                                               |               |                                              |                 |                                                |                  |                                                 |
| <b>Design Access:</b>  | Not Available                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                 |                                                                                                       |                |                                               |               |                                              |                 |                                                |                  |                                                 |
| <b>Runtime Access:</b> | Read-only                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                 |                                                                                                       |                |                                               |               |                                              |                 |                                                |                  |                                                 |
| <b>See Also:</b>       | RegIndex Property, AddRegion Method                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                 |                                                                                                       |                |                                               |               |                                              |                 |                                                |                  |                                                 |
| <b>Comments:</b>       | When the <b>OCRRegion</b> method is executed to begin recognition and no <b>RegionSource</b> is specified, the TextBridge control's own region definition properties are applied. The <b>RegCount</b> property reports the number of defined regions and the following additional properties are used to define those regions: <table> <tr> <td><b>RegIndex</b></td><td>May be set to any integer from 1 to <b>RegCount</b>. When set, the following properties are updated.</td></tr> <tr> <td><b>RegLeft</b></td><td>Image pixel coordinate of left edge of region</td></tr> <tr> <td><b>RegTop</b></td><td>Image pixel coordinate of top edge of region</td></tr> <tr> <td><b>RegRight</b></td><td>Image pixel coordinate of right edge of region</td></tr> <tr> <td><b>RegBottom</b></td><td>Image pixel coordinate of bottom edge of region</td></tr> </table> | <b>RegIndex</b> | May be set to any integer from 1 to <b>RegCount</b> . When set, the following properties are updated. | <b>RegLeft</b> | Image pixel coordinate of left edge of region | <b>RegTop</b> | Image pixel coordinate of top edge of region | <b>RegRight</b> | Image pixel coordinate of right edge of region | <b>RegBottom</b> | Image pixel coordinate of bottom edge of region |
| <b>RegIndex</b>        | May be set to any integer from 1 to <b>RegCount</b> . When set, the following properties are updated.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                 |                                                                                                       |                |                                               |               |                                              |                 |                                                |                  |                                                 |
| <b>RegLeft</b>         | Image pixel coordinate of left edge of region                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                 |                                                                                                       |                |                                               |               |                                              |                 |                                                |                  |                                                 |
| <b>RegTop</b>          | Image pixel coordinate of top edge of region                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                 |                                                                                                       |                |                                               |               |                                              |                 |                                                |                  |                                                 |
| <b>RegRight</b>        | Image pixel coordinate of right edge of region                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                 |                                                                                                       |                |                                               |               |                                              |                 |                                                |                  |                                                 |
| <b>RegBottom</b>       | Image pixel coordinate of bottom edge of region                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                 |                                                                                                       |                |                                               |               |                                              |                 |                                                |                  |                                                 |

### ***RegIndex Property***

|                        |                                                                                                                                                                                                       |
|------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Definition:</b>     | Specifies the active region for the reporting and setting of region coordinates.                                                                                                                      |
| <b>Data Type:</b>      | Long                                                                                                                                                                                                  |
| <b>Design Access:</b>  | Not Available                                                                                                                                                                                         |
| <b>Runtime Access:</b> | Read-only                                                                                                                                                                                             |
| <b>See Also:</b>       | RegCount Property, RegBottom Property, RegLeft Property, RegRight Property, RegTop Property                                                                                                           |
| <b>Comments:</b>       | When the <b>OCRRegion</b> method is executed to begin recognition and no <b>RegionSource</b> is specified, the TextBridge control's own region definition properties are applied. The <b>RegCount</b> |

property reports the number of defined regions and the following additional properties are used to define those regions:

|                  |                                                                                                       |
|------------------|-------------------------------------------------------------------------------------------------------|
| <b>RegIndex</b>  | May be set to any integer from 1 to <b>RegCount</b> . When set, the following properties are updated. |
| <b>RegLeft</b>   | Image pixel coordinate of left edge of region                                                         |
| <b>RegTop</b>    | Image pixel coordinate of top edge of region                                                          |
| <b>RegRight</b>  | Image pixel coordinate of right edge of region                                                        |
| <b>RegBottom</b> | Image pixel coordinate of bottom edge of region                                                       |

## ***RegionSource Property***

|                        |                                                                                                                                                                                                                                           |
|------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Definition:</b>     | Specifies the ImageBASIC control that will supply region coordinates for recognition when the <b>OCRRegion</b> method is executed.                                                                                                        |
| <b>Data Type:</b>      | String                                                                                                                                                                                                                                    |
| <b>Design Access:</b>  | Read/Write                                                                                                                                                                                                                                |
| <b>Runtime Access:</b> | Read/Write                                                                                                                                                                                                                                |
| <b>See Also:</b>       | ImageDataSource Property, OCRRegion Method                                                                                                                                                                                                |
| <b>Comments:</b>       | As of this writing, the primary control used for region definition is the Display control. The Working Region that is defined in that control is the image region that will be retrieved when <b>RegionSource</b> specifies that control. |

As an alternative to the **RegionSource** property, the following properties of the TextBridge control can be set to define the region for OCR:

RegBottom  
RegLeft  
RegRight  
RegTop

These properties may be used both when the image is received through the **ImageDataSource** property and when it is read directly from file. Refer to the 'InputFrom Property' on page 40 for details on selecting the source of image data.

## ***RegLeft Property***

|                        |                                                               |
|------------------------|---------------------------------------------------------------|
| <b>Definition:</b>     | Specifies the left edge of the region to OCR in image pixels. |
| <b>Data Type:</b>      | Long                                                          |
| <b>Design Access:</b>  | Read/Write                                                    |
| <b>Runtime Access:</b> | Read/Write                                                    |
| <b>See Also:</b>       | OCRRegion Method, RegIndex Property                           |

**Comments:** When the **OCRRegion** method is executed to begin recognition and no **RegionSource** is specified, the TextBridge control's own region definition properties are applied. The **RegCount** property reports the number of defined regions and the following additional properties are used to define those regions:

**RegIndex** May be set to any integer from 1 to **RegCount**. When set, the following properties are updated.

**RegLeft** Image pixel coordinate of left edge of region

**RegTop** Image pixel coordinate of top edge of region

**RegRight** Image pixel coordinate of right edge of region

**RegBottom** Image pixel coordinate of bottom edge of region

### ***RegRight Property***

**Definition:** Specifies the width of the region to OCR in image pixels.

**Data Type:** Long

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**See Also:** OCRRegion Method, RegIndex Property

**Comments:** When the **OCRRegion** method is executed to begin recognition and no **RegionSource** is specified, the TextBridge control's own region definition properties are applied. The **RegCount** property reports the number of defined regions and the following additional properties are used to define those regions:

**RegIndex** May be set to any integer from 1 to **RegCount**. When set, the following properties are updated.

**RegLeft** Image pixel coordinate of left edge of region

**RegTop** Image pixel coordinate of top edge of region

**RegRight** Image pixel coordinate of right edge of region

**RegBottom** Image pixel coordinate of bottom edge of region

### ***RegTop Property***

**Definition:** Specifies the top edge of the region to OCR in image pixels.

**Data Type:** Long

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**See Also:** RegLeft Property, RegBottom Property, RegionSource Property

**Comments:** When the **OCRRegion** method is executed to begin recognition and no **RegionSource** is specified, the TextBridge control's own region definition properties are applied. The **RegCount**

property reports the number of defined regions and the following additional properties are used to define those regions:

|                  |                                                                                                       |
|------------------|-------------------------------------------------------------------------------------------------------|
| <b>RegIndex</b>  | May be set to any integer from 1 to <b>RegCount</b> . When set, the following properties are updated. |
| <b>RegLeft</b>   | Image pixel coordinate of left edge of region                                                         |
| <b>RegTop</b>    | Image pixel coordinate of top edge of region                                                          |
| <b>RegRight</b>  | Image pixel coordinate of right edge of region                                                        |
| <b>RegBottom</b> | Image pixel coordinate of bottom edge of region                                                       |

### ***Result Property***

|                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Definition:</b>     | If <b>OutputTo</b> specifies <i>0--String</i> , reports the final recognition results after a successful OCR attempt.                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| <b>Data Type:</b>      | String                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <b>Design Access:</b>  | Not Available                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| <b>Runtime Access:</b> | Read-only                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <b>See Also:</b>       | OutputToProperty                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| <b>Comments:</b>       | The string that is reported in this property will be the final recognition results. The string will be formatted as specified in the <b>OutputFormat</b> property.<br><br>Valid only when <b>OutputTo</b> is set to <i>1--String</i> . After recognition is complete, this property will report the total number of regions that were processed. When the <b>ResultIndex</b> property is set to any integer value between 1 (one) and <b>ResultCount</b> inclusive, the <b>Result</b> property is populated with the output string for the specified region. |

### ***ResultCount Property***

|                        |                                                                                                                                                                                                                                                                                                                                                                                    |
|------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Definition:</b>     | Reports the number of processed regions. In this revision of the TextBridge control, will always report 1 (one).                                                                                                                                                                                                                                                                   |
| <b>Data Type:</b>      | Long                                                                                                                                                                                                                                                                                                                                                                               |
| <b>Design Access:</b>  | Not Available                                                                                                                                                                                                                                                                                                                                                                      |
| <b>Runtime Access:</b> | Read-only                                                                                                                                                                                                                                                                                                                                                                          |
| <b>See Also:</b>       | ResultIndexProperty, ResultProperty                                                                                                                                                                                                                                                                                                                                                |
| <b>Comments:</b>       | Valid only when <b>OutputTo</b> is set to <i>1--String</i> . After recognition is complete, this property will report the total number of regions that were processed. When the <b>ResultIndex</b> property is set to any integer value between 1 (one) and <b>ResultCount</b> inclusive, the <b>Result</b> property is populated with the output string for the specified region. |

### ***ResultIndex Property***

|                        |                                                                                                                                                                                                                                                                                                                                                                                                      |
|------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Definition:</b>     | Specifies the region whose OCR results will be reported in the <b>Result</b> property. In this revision of the TextBridge control, the only valid value is 1 (one).                                                                                                                                                                                                                                  |
| <b>Data Type:</b>      | Long                                                                                                                                                                                                                                                                                                                                                                                                 |
| <b>Design Access:</b>  | Not Available                                                                                                                                                                                                                                                                                                                                                                                        |
| <b>Runtime Access:</b> | Read/Write                                                                                                                                                                                                                                                                                                                                                                                           |
| <b>See Also:</b>       | ResultCountProperty, ResultProperty                                                                                                                                                                                                                                                                                                                                                                  |
| <b>Comments:</b>       | Valid only when <b>OutputTo</b> is set to <i>1--String</i> . After recognition is complete, the <b>ResultCount</b> property will report the total number of regions that were processed. When the <b>ResultIndex</b> property is set to any integer value between 1 (one) and <b>ResultCount</b> inclusive, the <b>Result</b> property is populated with the output string for the specified region. |

### ***SaveTraining Property***

|                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Definition:</b>     | If True, any learning performed either by the control or through the <b>Verify</b> event will be recorded in the <b>TrainFileName</b> file.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| <b>Data Type:</b>      | Boolean                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <b>Design Access:</b>  | Read/Write                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <b>Runtime Access:</b> | Read/Write                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <b>See Also:</b>       | TrainFileNameProperty, VerifyEvent                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <b>Comments:</b>       | <p>By default, if a <b>TrainFileName</b> is specified, the TextBridge control will write to that file each time a recognition is performed. It is through this process that the engine can learn on its own, becoming slightly more accurate over time, even without operator intervention and training. In some circumstances, the time and resources required to perform this learning may interfere with recognition.</p> <p>If the <b>SaveTraining</b> property is False, the learning file will not be modified by the control. Only the training already saved to the file will be applied. For more details on the training process, refer to 'Verification and Training' on page 26.</p> |

### ***TrainFileName Property***

|                        |                                                                                        |
|------------------------|----------------------------------------------------------------------------------------|
| <b>Definition:</b>     | Specifies the path and file name of the learning file to use for the next recognition. |
| <b>Data Type:</b>      | String                                                                                 |
| <b>Design Access:</b>  | Read/Write                                                                             |
| <b>Runtime Access:</b> | Read/Write                                                                             |

**See Also:** VerifyProperty

**Comments:** If LoadTraining is True, the control reads all verification information from this file before beginning recognition of each page and uses it to correct for uncertain characters.

If SaveTraining is True, all new learning information input by the user is written to this file at the end of every successful page. If the same **TrainFileName** is used in all your OCR attempts, TextBridge will continue to expand its knowledge base.

Learning information that will improve OCR recognition can be added to this file in two ways:

- 1) Any input from the **Verify** event will be recorded here
- 2) The OCR engine continually writes additional information here without any user input

### ***ThresholdAccept Property***

**Definition:** Specifies the threshold below which a character will be marked as uncertain.

**Data Type:** Integer

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**Possible Values:** 0 - 999, inclusive

**See Also:** VerifyProperty, ThresholdRejectProperty

**Comments:** As each character is recognized in an OCR attempt, the OCR engine assigns a confidence level between 0 and 999. The uncertain characters can be identified in the XDOC **OutputFormat** and during verification as described in "Verification and Training" on page 26.

### ***ThresholdReject Property***

**Definition:** Specifies the lower limit of confidence that a character must reach to be reported as questionable rather than noise.

**Data Type:** Integer

**Design Access:** Read/Write

**Runtime Access:** Read/Write

**Possible Values:** 0 - 999, inclusive

**See Also:** ThresholdAcceptProperty

**Comments:** The value of this property is compared to a confidence level during the verification process. Any character with a confidence level below the **ThresholdReject** value is replaced

by the noise character (~) in the output string. See "Verification and Training" on page 26 for a complete discussion.

### ***ThresholdVerify Property***

- Definition:** Specifies the minimum confidence that characters in the **return** string must reach; if any character does not reach this threshold, the **Verify** event will occur for each character if the **Verify** property is True.
- Data Type:** Integer
- Design Access:** Read/Write
- Runtime Access:** Read/Write
- Possible Values:** 0 - 999, inclusive
- See Also:** Verify Property, ThresholdRejectProperty
- Comments:** When the Verification Dialog is displayed, all characters that fall below **ThresholdAccept** will be highlighted for verification or correction. If the **TrainFileName** is set to a valid file name and the **SaveTraining** property is True, all verification information will be stored in the specified file for future use.

### ***Verify Event***

- Definition:** Occurs for each character or group that falls below the confidence level set in **ThresholdVerify** property. Enabled by setting the **Verify** property to True.
- Parameters:**
- BestGuess* String of the engine's interpretation of the image
  - SuspectStart* Long position of the character that triggered the event in the *BestGuess* string
  - SuspectLength* Long number of characters in the *BestGuess* string that are in the suspect group
  - Top* Long image pixel coordinate of the top of the bounding rectangle of the suspect group
  - Bottom* Long image pixel coordinate of the bottom of the bounding rectangle of the suspect group
  - Left* Long image pixel coordinate of the left edge of the bounding rectangle of the suspect group
  - Right* Long image pixel coordinate of the right edge of the bounding rectangle of the suspect group
  - Correction* Can be set to the correct interpretation of the suspect group
  - Action* Set to one of the options listed below to instruct the engine how to continue processing or learning

*Cancel* If set to True, this event is ended without any changes

**Comments:**

The term *group* refers to any collection of black pixels which is contiguous. Normally, this refers to any single character. Sometimes, however, several characters are interpreted as a single group, and on other occasions one letter is broken into more than one part. In instances such as these, the options of joining adjacent groups and allowing the user to type in the correct interpretation of the entire string of characters.

When the **Verify** event occurs the TextBridge control can be a source of image data for a display control. For example, a TMSSequoia Display control may be used to show the region of the image that contains the uncertain character or word:

```
TMSDisp2.ImageDataSource = TextBridge1.Link
```

Following are the enumerated options for the *Action* parameter.

- 0 Accept Group
- 1 Replace Group
- 2 Join Right
- 3 Join Left
- 4 Accept and Move Right
- 5 Accept and Move Left
- 6 Undo
- 7 Accept Word
- 8 Accept Word No Train

0--*Accept Group* accepts the character(s) in the *BestGuess* parameter as they appear. Any entry made in the *Correction* parameter will be ignored.

Recognition will now continue, and the **Verify** event will occur again for the next character which falls below

**ThresholdVerify**

1--*Replace Group* causes the engine to replace the group that triggered this event with the string to which the *Correction* parameter is set.

This correction will be saved for future recognition events if a valid file is specified in the **TrainFileName** property and **SaveTraining** is True.

2--*Join Right* causes the current group to be joined to the group immediately to its right. This feature will be used when a single letter is split in the image and is considered to be two characters by the recognition engine.



The **Verify** event will occur again for the newly joined group. No verification information or training is created by this action.

3--*Join Left* causes the current group to be joined to the group immediately to its left. This feature will be used when a single letter is split in the image and is considered to the two characters by the recognition engine.

The **Verify** event will occur again for the newly joined group. No verification information or training is created by this action.

4--*Accept Right* causes the string to be accepted as originally presented in the *BestGuess* parameter and triggers the **Verify** event again for the group to the right of the current group.

5--*Accept Left* causes the string to be accepted as originally presented in the *BestGuess* parameter and triggers the **Verify** event again for the group to the left of the current group.

6--*Undo* causes the most recently entered information to be undone. For example, if in the previous **Verify** event, the *Action* parameter was set to 5--*Replace Group* and the training file was updated based on that entry, selecting the *Undo Action* will remove that entry from the training file and set the output string back to the engine's original interpretation.

7--*Accept Word* causes the string to be accepted as originally presented in the *BestGuess* parameter.

This confirmation will be saved for future recognition events if a valid file is specified in the **TrainFileName** property and **SaveTraining** is True.

8--*Accept Word No Train* causes the string to be accepted as originally presented in the *BestGuess* parameter.

This confirmation will not be saved for future recognition events even if a valid file is specified in the **TrainFileName** property and **SaveTraining** is True.

## **Verify Property**

|                        |                                                                                                                                                                 |
|------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Definition:</b>     | If True, the <b>Verify</b> event will occur once for each character or group which falls below the confidence level set in the <b>ThresholdVerify</b> property. |
| <b>Data Type:</b>      | Boolean                                                                                                                                                         |
| <b>Design Access:</b>  | Read/Write                                                                                                                                                      |
| <b>Runtime Access:</b> | Read/Write                                                                                                                                                      |
| <b>See Also:</b>       | "Verification and Training" on page 26                                                                                                                          |
| <b>Comments:</b>       | The <b>Verify</b> event will occur once for any character that falls below <b>ThresholdVerify</b> however, the engine will specifically                         |

request correction only for characters that fall below the minimum confidence level specified in the **ThresholdAccept** property.

### ***YieldLevel Property***

- Definition:** If set to a non-zero value, the control will surrender that number of milliseconds of CPU time during each second of an OCR attempt.
- Data Type:** Integer
- Design Access:** Read/Write
- Runtime Access:** Read/Write
- Comments:** The OCR of a single region generally does not require so much time that parallel processes cannot wait for completion. Unless circumstances dictate otherwise, this property should be left at its default of 0.





# Appendix A : Supported Formats and DLLs

---

## Output Formats and Required Conversion Libraries

The following table lists the file format conversion libraries required for each of the output file formats. In order for the TextBridge control to write a particular file format with its OCR results, the DLL listed in the table must be available along with xisconv.dll (16-bit) or xiscnv32.dll (32-bit).

|                        |             |
|------------------------|-------------|
| AmiPro (All)           | w4w33t.dll  |
| ASCII (All)            | w4w01t.dll  |
| dBASE IV 1.0           | w4w770t.dll |
| DCA/RTF                | w4w15t.dll  |
| Excel (All)            | w4w21t.dll  |
| FrameMaker             | w4w42t.dll  |
| HTML (All)             | w4w108t.dll |
| Interleaf              | w4w46t.dll  |
| Lotus 1-2-3            | w4w20t.dll  |
| MultiMate Adv. II      | w4w10t.dll  |
| Postscript             | w4w40t.dll  |
| Professional Write 2.x | w4w08t.dll  |
| RTF (All)              | w4w19t.dll  |
| Samna Word IV          | w4w22t.dll  |
| Windows Write          | w4w43t.dll  |
| Word for Windows (All) | w4w19t.dll  |
| WordPerfect Windows    | w4w07t.dll  |
| WordPerfect 4.2        | w4w06t.dll  |
| WordPerfect Win 6.x    | w4w48t.dll  |
| WordStar Windows 1.x   | w4w37t.dll  |



# Index

## A

- AboutBox Method 31
- Action Parameter 60
- Active Property 31

## B

- Bottom Parameter 60

## C

- Cancel Parameter 33, 60
- Columns
  - OptAutoSegment Property 43
- Columns in Originals 20
- Completion Event 33
- Completion Event Parameters
  - Cancel 33
  - Percent 33
- Correction Parameter 60

## D

- Data Input
  - ImageDataSource Property 1
- Database, Creation of Licensing 3
- Database, Licensing 2
- DebugPath Property 33
- Degraded Images
  - OptDoRight Property 45
- Degraded Originals 22
- Deskewing
  - OptAutoDeskew Property 42
- Dictionaries (Lexicons) 40
- Distorted Originals 21
- Dot Matrix Originals 21
- Dot Matrix Text
  - OptDotDetect Property 45
  - OptDotMatrix Property 45

## E

- Error Event 34
- Event Parameters
  - Action 60
  - BestGuess 60
  - Bottom 60
  - Cancel 33, 60
  - Correction 60
  - Left 60

- Percent 33
- Right 60
- SuspectLength 60
- SuspectStart 60
- Top 60

## Events

- Completion 33
- Error 34
- ImageDataChanged 2
- Verify 25, 60

## F

- Fax Documents 21
- Faxed Text
  - ICRMode Property 34
  - OptAutoFax Property 42
  - OptAutoSetDegrade Property 44
  - OptFax Property 46
- Field Type Definition 38
- Field Types 16, 19
- File Input
  - PageIndex Property 53
- File Names, Lexicons 40
- File Output
  - OutputFileAppend Property 50
  - OutputFileName Property 50
  - OutputFormat Property 51
  - OutputTo Property 52
- File to File
  - See InputFrom Property 6
  - See OutputTo Property 8
- File-to-File Processing
  - InputFileName Property 35
  - InputFrom Property 36
  - OutputTo Property 52
  - PageIndex Property 53
- Foreign Language Recognition 14, 37

## H

- Halftone Originals 22
- Halftone Shading Removal
  - OptRemoveHalftone Property 47

## I

- ICRMode Property 34
- Image Transfer
  - Link Property 40
- ImageDataChanged Event 2
- ImageDataSource Property 1, 35
- ImageResolution Property 36

- InputFileName Property 35
- InputFrom Property 36
- Inverse Originals 23
- Inverse Text
  - Invert Property 37
  - TextOptInvert Property 46
- Invert Property 37
- L**
- Language Lexicons 14, 17
- Language Property 16, 37
- Language Support 14
- LCM, Using 3
- Left Parameter 60
- LexicalClass Property 16, 38
- LexicalMode Property 16, 39
- LexiconFileName Property 16, 40
- Lexicons
  - LexicalClass 38
  - LexiconFileName Property 40
- License Storage 2
- License Verification 3
- Licensing
  - Active Property 31
- Limiting OCR Results 38
- Link Property 40
- Linking
  - ImageDataSource Property 35
- Linking Controls
  - ImageDataChanged Event 2
  - ImageDataSource Property 1
- Load Time Improvement
  - Active Property 31
- LoadTraining Property 40
- Locating Lexicons 40
- Low Quality Originals 21
- Low Resolution
  - ICRMode Property 34
  - OptAutoFax Property 42
  - OptAutoSetDegrade Property 44
  - OptFax Property 46
- M**
- Methods
  - AboutBox 31
  - OCRPage 41
  - OCRRegion 41
- Multiple Region Definition
  - RegCount Property 54
  - RegIndex Property 54

- Multiple-page Image Files 53

## N

- Newspaper
  - OptNewspaper Property 47

## O

- OCR Results
  - Result Property 57
- OCRPage Method 41
- OCRRegion Method 41
- OptAutoDeskew Property 42
- OptAutoFax Property 42
- OptAutoPortrait Property 43
- OptAutoSegment Property 43
- OptAutoSetDegrade Property 44
- OptAutoUpright Property 44
- OptDoRight Property 45
- OptDotDetect Property 45
- OptDotMatrix Property 45
- OptFax Property 46
- OptFullOrientation Property 46
- OptInvert Property 46
- OptNewspaper Property 47
- OptRemoveHalftone Property 47
- OptRotation Property 47
- OptXDocCharConf Property 48
- OptXDocLexicalClass Property 48
- OptXDocWordBox Property 49
- OptXDocWordConf Property 49
- Orientation
  - OptAutoPortrait Property 43
  - OptAutoUpright Property 44
  - OptFullOrientation Property 46
  - OptRotation Property 47
  - Orientation Property 49
- Orientation Property 49
- OutputFileAppend Property 50
- OutputFileName Property 50
- OutputFormat Property 51
- OutputTo Property 52

## P

- PageIndex Property 53
- Parameter 60
- Percent Parameter 33
- Properties
  - Active 31
  - DebugPath 33
  - ICRMode 34



- ImageDataSource 1, 35
- ImageResolution 36
- InputFileName 35
- InputFrom 36
- Invert 37
- Language 14, 16, 37
- LexicalClass 16
- LexicalMode 16, 39
- LexiconFileName 16
- Link 40
- LoadTraining 40
- OptAutoDeskew 42
- OptAutoFax 42
- OptAutoPortrait 43
- OptAutoSegment 43
- OptAutoSetDegrade 44
- OptAutoUpright 44
- OptDoRight 45
- OptDotDetect 45
- OptDotMatrix 45
- OptFax 46
- OptFullOrientation 46
- OptInvert 46
- OptNewspaper 47
- OptRemoveHalftone 47
- OptRotation 47
- OptXDocCharConf 48
- OptXDocLexicalClass 48
- OptXDocWordBox 49
- OptXDocWordConf 49
- Orientation 49
- OutputFileAppend 50
- OutputFileName 50
- OutputFormat 51
- OutputTo 52
- PageIndex 53
- RecognizeCount 53
- RegBottom 53
- RegCount 54
- RegIndex 54
- RegionSource 55
- RegLeft 55
- RegRight 56
- RegTop 56
- Result 57
- ResultCount 57
- ResultIndex 58
- SaveTraining 26, 58
- ThresholdAccept 25, 39, 59
- ThresholdReject 25, 59

- ThresholdVerify 25, 60
- TrainFileName 26, 58
- Verify 25
- YieldLevel 63

## R

- Recognition
  - OCRPage Method 41
  - OCRRegion Method 41
  - RegBottom Property 53
  - RegCount Property 54
  - RegIndex Property 54
  - RegionSource Property 55
  - RegLeft Property 55
  - RegRight Property 56
  - RegTop Property 56
- Recognition Output 52
  - Result Property 57
- RecognizeCount Property 53
- RegBottom Property 53
- RegCount Property 54
- RegIndex Property 54
- RegionSource Property 55
- RegLeft Property 55
- RegRight Property 56
- RegTop Property 56
- Releasing License Tokens 4
- Result Property 57
- ResultCount Property 57
- ResultIndex Property 58
- Right Parameter 60
- Rotated Images
  - OptAutoPortrait Property 43
  - OptAutoUpright Property 44
  - OptFullOrientation Property 46
  - OptRotation Property 47
  - Orientation Property 49
- Rotated Originals 23

## S

- SaveTraining Property 26, 58
- Segmentation
  - OptAutoSegment Property 43
- Segmenting Cols 20
- Shading Removal
  - OptRemoveHalftone Property 47
- Skewed Originals 24
- Skewed Text
  - OptAutoDeskew Property 42
- Source of Image Data 36

SuspectLength Parameter 60  
SuspectStart Parameter 60

## T

ThresholdAccept Property 25, 39, 59  
ThresholdReject Property 25, 59  
ThresholdVerify Property 25, 60  
Timing of License Locking/Unlocking 3, 4  
Timing of Licensing Verification  
    Active Property 31  
Top Parameter 60  
TrainFileName Property 26, 58  
Training 26  
    LoadTraining Property 40  
    SaveTraining Property 58

## U

User-Defined Lexicons 15, 17  
Using License Tokens 3

## V

Verification File Saving  
    LoadTraining Property 40  
    SaveTraining Property 58  
Verify Event 25, 60  
Verify Event Parameters  
    Action 60  
    Bottom 60  
    Cancel 60  
    Correction 60  
    Left 60  
    Right 60  
    SuspectLength 60  
    SuspectStart 60  
    Top 60  
Verify Property 25, 62

## W

White-on-Black Text  
    Invert Property 37  
    OptInvert Property 46  
Word Count  
    RecognizeCount Property 53

## X

XDoc Output Options  
    OptXDocCharConf Property 48  
    OptXDocLexicalClass Property 48  
    OptXDocWordBox Property 49  
    OptXDocWordConf Property 49

## Y

YieldLevel Property 63