



GLOSARIO DE TÉRMINOS

En este glosario se definen términos y conceptos usados en este manual. Para aquellos lectores sin experiencia previa en los conceptos de exploración y de reconocimiento de caracteres, este glosario puede serles útil no solo como referencia sino también en la comprensión de la tecnología de reconocimiento de caracteres.

Algunas definiciones contienen términos en **negrita**. Esto significa que estos términos también se definen en el glosario.

A **aplicación**— Programa de software que permite al usuario efectuar una tarea o varias. También se refiere a veces, al uso (esto es, la “aplicación”) de un programa de software.

ASCII— Acrónimo de American Standard Code for Information Interchange. ASCII contiene los códigos para 128 caracteres, alfanuméricos, símbolos y de control. Existen también otros grupos de caracteres ASCII extendidos que generalmente asignan otros 128 códigos para aquellos caracteres acentuados y aquellos símbolos no incluidos en los 128 primeros.

B **brillo del escáner**— Una asignación que determina la intensidad de luz que el escáner proyecta en la página que está siendo explorada para aclarar u oscurecer la imagen final. Al ajustarse el brillo, puede manipularse la exactitud del reconocimiento (por ejemplo, aclarar una página cuyos caracteres están muy juntos para mejorar el reconocimiento). En TextBridge puede especificar el brillo del escáner en el cuadro ‘Parámetros del escáner’.

C **casilla de edición**—Convención de la interfaz Windows que consiste de un campo rectangular en un cuadro u otra zona de la pantalla, en la que el usuario escribe texto o en la que existe texto que puede editarse. En Windows, una casilla de edición respalda las formas estándares que el usuario utiliza para editar la información. Por ejemplo, en el cuadro ‘Guardar como’ de una aplicación Windows la zona en la que se escribe el nombre del archivo de salida es una casilla de edición estándar.

CCITT—Acrónimo de Comité Consultivo Internacional de Telefonía y Telegrafía, organización internacional de estándares, la cual ha creado entre otras cosas, los estándares de compresión para los datos digitales. Los archivos TIFF almacenados en los estándares de compresión CCITT Group 3 y Group 4 puede ser reconocidos por TextBridge.

colores omitidos—Color o rango de colores con los que el escáner tiene dificultad en detectar en la página que está explorando. Esto es debido a la fuente de luz propia del escáner. Una fuente de luz amarilla en el escáner por ejemplo, representará un problema al detectar los colores en el espectro del amarillo al verde claro en la página explorada.

configuración del escáner—Un elemento del grupo programas TextBridge OCR en Windows. Este programa está diseñado para permitir cargar un controlador más sofisticado para que TextBridge pueda ejecutarse con el escáner. Véase también **ISIS** y **TWAIN**.

controlador de memoria expandida—Programa que hace que parte de la memoria extendida sea tratada como memoria adicional expandida para que los programas que requieren más de los 360K de memoria expandida disponibles en una máquina DOS puedan ejecutarse.

controlador del escáner—Programa escrito como interfaz entre una aplicación de software y un escáner. El controlador de escáner envía solicitudes desde la aplicación al escáner en un lenguaje que el escáner puede entender.

conversión—Módulo de software que acepta texto en un formato (el **formato de entrada**) y lo procesa a otro formato (el **formato de salida**). En TextBridge, el texto reconocido en el formato interno puede por ejemplo, convertirse a WordPerfect (o en cualquier otro de los formatos respaldados).

cuadro de diálogo—En Microsoft Windows, una categoría de pantalla de la interfaz del usuario que requiere interactuar (“diálogo”) con el usuario de la aplicación. TextBridge muestra un cuadro principal desde el cual se pueden definir e iniciar trabajos de reconocimiento.

D **delimitadores**—Los cuadrados oscuros que normalmente están en las cuatro esquinas de un rectángulo en un paquete de dibujo u otra aplicación que permita ajustar el tamaño de un rectángulo. En TextBridge, puede dibujarse una **zona** rectangular en una imagen de página mostrada primero para definir el área de la página a ser reconocida. Mediante los delimitadores puede ajustarse el tamaño de la zona.

directorio de trabajo—En Windows, cuando se instala una aplicación o después de haberla instalado, puede designarse un directorio en el sistema de archivos DOS como directorio de trabajo. Esto se efectúa mediante el ‘Administrador de programas’ seleccionando el comando ‘Propiedades’ del menú ‘Archivo’. Para TextBridge, el directorio de trabajo es un subdirectorio BIN en el directorio de instalación que se ha especificado durante el proceso de instalación. El programa de instalación de TextBridge elige como opción predeterminada el directorio de trabajo C:\TXBRIDGE\BIN, aunque puede seleccionarse cualquier directorio en el sistema de archivos DOS.

disco RAM—Parte de la memoria extendida de la computadora asignada como si fuera un disco duro para poder almacenar archivos temporalmente.

Dynamic Data Exchange (DDE)—Intercambio dinámico de datos. En el entorno Microsoft Windows, un estándar para la compartición de datos entre aplicaciones Windows. Por ejemplo, con la macro adecuada, un programa de tratamiento de textos como Microsoft Word para Windows puede indicar a TextBridge que explore y reconozca texto e importar el texto a un documento Word desde su propio sistema de menús.

F formato de galerada—Formato de una sola columna en el que TextBridge genera el texto reconocido de documentos con varias columnas.

formato de texto—Tratamiento de texto, hoja electrónica u otro formato de archivo al cual puede convertirse y generarse el texto reconocido. TextBridge respalda la salida de texto reconocido en los siguientes formatos:

Ami Pro (2.0, 3.0)	PostScript
ASCII (Standard, Smart, Stripped)	Prof Write (2.0, 2.2)
dBase IV	RTF (Rich Text Format de Microsoft)
DCA/RFT	Samna Word IV
DisplayWrite 5	Windows Write
Excel (Mac, 3.0, 4.0)	Word para Windows 2.0
FrameMaker	WordPerfect (4.2, 5.1)
Interleaf	WordStar
Lotus 1-2-3	
Multimate Advantage	

fuentes de entrada—Origen de las imágenes de página que están siendo reconocidas: ya sea el escáner o un archivo TIFF.

- I idioma de reconocimiento**—Idioma principal (por ejemplo, español, inglés) del documento. En TextBridge puede especificarse que el documento está en un idioma o en varios. TextBridge cargará el **paquete de idioma** correcto antes de empezar el reconocimiento.

imagen de fax—Representación de una página en formato de datos binarios (normalmente 200x100 ó 200x200 ppp de resolución) transmitidos mediante una máquina facsímil (fax) o una tarjeta módem fax. Las computadoras con tarjetas módem fax pueden recibir una imagen de fax y almacenarla en línea como un archivo TIFF. TextBridge puede abrir y reconocer el texto de una imagen de fax en línea almacenada en formato TIFF.

imagen de página—Fotografía binaria (en blanco y negro) de una página almacenada en la memoria de una computadora o en disco. Las imágenes de página pueden ser exploradas o leídas desde un archivo TIFF y enviadas a TextBridge para el reconocimiento óptico de caracteres (OCR).

imagen TIFF—Representación binaria de una página o gráfico almacenado en formato Tagged-Image File Format, formato de archivo de imagen estándar en la industria. TextBridge puede reconocer texto de imágenes de página almacenados en distintas variaciones de TIFF, como:

- TIFF Sin comprimir (Cabecera Intel)
- TIFF CCITT-3 (Cabecera Intel)
- TIFF CCITT-4 (Cabecera Intel)

- TIFF Sin comprimir (Cabecera Motorola)
- TIFF CCITT-3 (Cabecera Motorola)
- TIFF CCITT-4 (Cabecera Motorola)

- TIFF (Cabecera Intel FAXability™)

Al seleccionar '**Guardar imagen de página**' en el cuadro principal, TextBridge guarda automáticamente las imágenes exploradas de página a archivos TIFF CCITT-3 Intel.

impresora OCR—Aplicación de TextBridge diseñada para que funcione como una impresora desde cualquier programa de fax o de imágenes basado en Windows. En el programa principal, se visualiza un fax u otra imagen que contiene texto y el comando Imprimir se usa para enviarlos a la impresora OCR. El cuadro 'Guardar como' permite guardar el texto a un archivo en el formato deseado.

interfaz nativa—En controladores **TWAIN**, conjunto de visualizaciones de pantalla y controles de teclado que un controlador de origen TWAIN proporciona a los programas que respaldan el dispositivo TWAIN. Por ejemplo, TextBridge puede ejecutarse con escáneres que cumplen con el estándar TWAIN. Sin embargo, los controles para dicho escáner son proporcionados en la interfaz nativa y no en TextBridge.

ISIS—Acrónimo de **Image and Scanner Interface Standard** (estándar de la interfaz de imagen y escáner) desarrollado por Pixel Translations, Inc. ISIS es una interfaz de los programadores de aplicaciones usadas para el diseño y desarrollo de controladores de escáner. Pixel Translations y otros fabricantes de escáneres usan ISIS para desarrollar los controladores de escáner. TextBridge respalda la mayoría de escáneres compatibles con ISIS del mercado. Véase también **TWAIN**.

M **medios tonos**—Imagen compuesta por puntos negros de varios tamaños y separados de tal forma que representan distintas tonalidades de gris de una fotografía original o de un dibujo a color. En el cuadro 'Preferencias', puede especificarse que TextBridge ignore los medios tonos cuando se está efectuando un reconocimiento óptico de caracteres.

memoria virtual—Espacio en el disco duro asignado para simular la memoria de acceso aleatorio (RAM) en la PC. TextBridge, especialmente en sistema con sólo 4Mb de RAM, debe ser configurado con memoria virtual **permanente**, grupo de bloques de almacenamiento contiguos en el disco duro.

módem fax—Dispositivo externo o placa de circuito impresa que se conecta a la PC permitiendo recibir y transmitir datos de imagen digital a través de una línea de telecomunicaciones (teléfono). Con un módem fax conectando la PC con la línea telefónica, pueden recibirse y transmitirse imágenes de documento de y a la PC.

modo extendido—El modo más avanzado de los tres modos en los que puede ejecutarse Microsoft Windows. Los otros dos modos son real y estándar. TextBridge puede ejecutarse sólo en modo extendido.

Muestra—En TextBridge, una función que permite ver, acercar o alejar y definir una zona en una página antes de procesarla.

N nombre base—La parte del **nombre del documento** usado para identificar imágenes de página relacionadas que los nombres de archivos (TIFF) han creado al utilizarse la función ‘Guardar imágenes de página’ de TextBridge. Al escribir el nombre de un documento en el cuadro ‘Guardar imágenes de página como’, los primeros cuatro dígitos del nombre son usados como el nombre base de los archivos de imagen de página.

nombre del documento—Nombre del archivo que se ha introducido en los cuadros ‘Guardar imágenes de página como’ o ‘Guardar como’ en TextBridge. El nombre del documento es anexado automáticamente a una extensión de tres letras lo que indica el formato en el cual el archivo reconocido va a ser guardado.

O **orientación de página automática**—En TextBridge, la capacidad de corregir la rotación de la imagen de página antes de empezar el reconocimiento de una página. Por ejemplo, si el usuario estuviera explorando un documento con páginas en distintas orientaciones (por ejemplo, la mayoría de páginas en vertical, algunas en horizontal con tablas de gran tamaño), TextBridge puede efectuar una orientación automática en cada página antes de empezar el reconocimiento.

OS/2—Sistema operativo gráfico diseñado por International Business Machines Corporation para (IBM®) ejecutarse con computadoras personales basadas en Intel. OS/2 es un sistema operativo multiprocesador que puede ejecutar programas nativos y programas diseñados para Microsoft Windows y DOS.

P **palabra dudosa**—Palabra que está por debajo del nivel de precisión de TextBridge. Durante el reconocimiento, TextBridge asigna un valor de precisión a cada palabra. Si dicho valor está por debajo del nivel de precisión definido, y se está usando el '**Verificador**', TextBridge mostrará la palabra como dudosa.

paquete de idioma—Componente de TextBridge que permite a la aplicación efectuar un reconocimiento en un documento creado en un idioma en particular. En el cuadro Preferencias de TextBridge puede especificarse el idioma del documento a reconocerse: alemán, español, francés, inglés o italiano. TextBridge carga el paquete de idioma adecuado antes de empezar el reconocimiento. Véase también **idioma de reconocimiento**.

permanente—Véase **memoria virtual**.

píxel—Uno de los muchos puntos de los que está formada una imagen digital.

preferencias—En TextBridge, las asignaciones que pueden especificarse para controlar el proceso de reconocimiento.

programa TSR—Programa diseñado para cargarse en memoria automáticamente al arrancar el sistema o a permanecer en memoria una vez se ha salido del programa.

R **reconocimiento**—Proceso de TextBridge durante el cual una imagen de página (explorada o TIFF en línea) es analizada y se identifican los caracteres y palabras además de guardarse como datos de texto en memoria o en un archivo temporal en línea. Observe que el sistema de reconocimiento TextBridge no solo efectúa el reconocimiento sino que efectúa una segmentación, orientación (rotación), un análisis de formato y retiene los estilos de texto (negrita, itálica).

reconocimiento óptico de caracteres (OCR)—Tecnología por la que las imágenes binarias de formas de caracteres son analizadas e identificadas como caracteres y generadas en un flujo de datos de texto ya sea a la memoria de la computadora o a un archivo de computadora.

región—Bloque lógico de tipo en una imagen de página. TextBridge, mediante la función de **segmentación automática** localiza las regiones de texto en un documento de varias columnas y las genera en el orden correcto.

resolución—Nivel de detalle medido en **puntos por pulgada** (ppp) mediante el cual un escáner o máquina de fax puede introducir una imagen. TextBridge puede efectuar un reconocimiento óptico de caracteres en imágenes de página con las siguientes resoluciones (puntos por pulgada): 400x400, 400x200, 300x300, 200x200, 200x100 y 100x200.

RTF—Siglas de Rich Text Format, formato de texto desarrollado por Microsoft Corporation con códigos incrustados para describir tipos de letra, formatos, etc.

S **segmentación automática de página**—En TextBridge, la capacidad de discernir la disposición de una imagen de página y en reconocer y generar texto en el orden correcto. Por ejemplo, en un boletín informativo, en el que las columnas son desiguales en ancho y largo, TextBridge reconoce la disposición de la página y genera el texto en la secuencia correcta. En la ventana 'Preferencias' de TextBridge, puede activarse o desactivarse la segmentación automática de páginas.

sentencia DEVICE—En el archivo `config.sys`, una línea que identifica por ejemplo, el controlador del dispositivo de un escáner a las aplicaciones que necesite ejecutar el escáner.

suciedad—Marca en una página que puede ser reconocida erróneamente como uno o más caracteres durante el reconocimiento.

T **texto de salida**—Véase **formato de texto**.

texto enlazado—Capacidad de atravesar la base de datos de texto (en un sistema de ayuda en línea) de varias formas: seleccionando un tema de un índice; pasando de tema a tema hacia delante o hacia atrás; efectuando una búsqueda de una palabra clave; por contexto (haciendo clic en una palabra del texto para obtener la definición o bien otra pantalla que contiene información sobre dicha palabra). TextBridge usa el sistema de la ayuda de Microsoft Windows para su propio sistema de ayuda incorporado y basado en texto enlazado.

TWAIN—Interfaz estándar de imagen y escáner completa con un API para el desarrollo de interfaces a dispositivos de imagen (escáneres, máquinas de fax etc). TextBridge respalda cualquier escáner que cumpla con las normas TWAIN o cualquier otro dispositivo que se conecte a una PC y reproduzca imágenes binarias (en blanco y negro) en un tamaño y resolución respaldados.

V Verificador—Función que permite ver y, si es necesario, corregir palabra por palabra las decisiones de reconocimiento tomadas por TextBridge. Una ventana ‘Verificador’ similar a la ventana ‘Muestra’ visualiza la palabra reconocida y su imagen asociada en la página explorada. La palabra reconocida es realzada en una casilla de edición, permitiendo hacer correcciones de ser necesario.

W Windows—Interfaz de usuario gráfica (GUI) y un grupo de módulos relacionados desarrollados por Microsoft Corporation para usar en computadoras personales con sistema operativo DOS, versión 3.1 y posterior.

Z zona—En la ventana ‘Muestra’ de TextBridge, un reborde que puede dibujarse alrededor de una parte de la imagen de página visualizada en el área de página a procesarse.

zoom—En la ventana ‘Muestra’ de TextBridge, la capacidad de aumentar (“zoom para acercar”) una imagen de página a resolución completa y restablecerla (“zoom para alejar”) a una resolución que permite visualizar la imagen de página completa.