



GLOSSARIO DEI TERMINI

Il presente glossario definisce termini e concetti utilizzati nel manuale. Per gli inesperti di scansione e dei principi del riconoscimento dei caratteri, questo glossario può rivelarsi utile non solo per la consultazione, ma anche come manualetto della tecnologia di riconoscimento ottico dei caratteri.

Alcune definizioni riportate qui contengono dei termini in **grassetto**, ad indicare che questi termini sono definiti anche in altri punti del glossario.

A **Anteprima**—In TextBridge, indica la funzione che consente, prima dell'elaborazione, di visualizzare, ingrandire/ridurre una pagina o definirne una zona.

applicazione—Programma software che consente agli utenti di eseguire un compito o una serie di compiti. A volte indica anche l'uso (o "applicazione") di un programma software.

ASCII—American Standard Code for Information Interchange. Il codice di interscambio delle informazioni ASCII contiene codici per 128 caratteri di controllo, alfanumerici e simboli. Alcuni dei cosiddetti set estesi ASCII assegnano generalmente altri 128 codici per i caratteri accentati e per simboli supplementari non inclusi nei primi 128.

C **casella di modifica**—Convenzione di interfaccia Windows rappresentata da un campo rettangolare all'interno di una finestra di dialogo o un'altra area dell'interfaccia, in cui l'utente può digitare del testo o modificare quello esistente. In Windows, le informazioni contenute nella casella di modifica possono essere modificate in alcuni modi standard. Ad esempio, nella finestra di dialogo standard Salva con nome delle applicazioni Windows, l'area in cui viene digitato il nome del file di uscita è rappresentata da una casella di modifica standard.

CCITT—Abbreviazione di *Comité Consultatif International Téléphonique et Télégraphique* (Comitato consultivo internazionale per la telegrafia e la telefonia), un'organizzazione internazionale degli standard che ha creato, tra l'altro, gli standard di compressione per i dati digitalizzati. I file TIFF memorizzati negli standard di compressione Gruppo 3 e Gruppo 4 CCITT possono essere riconosciuti da TextBridge.

colore ignorato—Un colore o una gamma di colori che lo scanner individua con difficoltà sulla pagina digitalizzata. Di solito, è il prodotto della stessa sorgente luminosa dello scanner. Ad esempio, una sorgente luminosa gialla dello scanner avrà problemi nell'individuazione dei colori compresi tra il giallo e il verde chiaro.

conversione—Modulo di software che acquisisce il testo in un formato (il **formato di entrata**) e lo trasforma in un altro formato (il **formato di uscita**). In TextBridge, il testo riconosciuto nel suo formato interno può essere convertito, ad esempio, in formato WordPerfect o in uno dei vari formati supportati.

D **directory di lavoro**—In Windows, quando viene installata un'applicazione, o dopo questa operazione, è possibile specificare come directory di lavoro una directory esistente all'interno del sistema di file DOS. Per farlo, nel Program Manager si seleziona il comando Proprietà dal menu File. Per TextBridge, la directory di lavoro è una sottodirectory BIN della directory di installazione, e viene specificata al momento dell'installazione. Il programma di installazione di TextBridge seleziona la directory di lavoro predefinita C:\TXBRIDGE\BIN, anche se è possibile selezionare una directory qualsiasi da qualsiasi partizione del sistema di file DOS.

disco RAM—Parte della memoria estesa del computer che agisce come un vero e proprio disco fisso per la memorizzazione temporanea dei file.

driver della memoria espansa—Programma che fa apparire una parte della memoria estesa come memoria espansa supplementare, per rendere possibile l'esecuzione dei programmi che richiedono più di 360K di memoria espansa, di solito presenti nelle macchine DOS.

driver di scanner—Programma che agisce da interfaccia tra un'applicazione software e uno scanner. Il driver di scanner invia le richieste dall'applicazione allo scanner, utilizzando un linguaggio che lo scanner può comprendere.

Dynamic Data Exchange (DDE)—In ambiente Microsoft Windows, indica uno standard di condivisione dei dati tra le applicazioni Windows. Ad esempio, con la macro adeguata e al posto giusto, un elaboratore di testi come Microsoft Word per Windows può dare istruzioni a TextBridge di digitalizzare e riconoscere il testo e di importarlo in un documento Word aperto, dall'interno del suo sistema a menu.

E elemento estraneo—Segno presente sulla pagina, che durante l'OCR può essere scambiato per uno o più caratteri.

F fax modem—Periferica esterna o scheda circuito che viene collegata al PC, consentendo di ricevere e trasmettere dati di immagini digitalizzate attraverso la linea delle telecomunicazioni (telefono). Collegando alla linea telefonica il PC con la scheda fax modem, il PC può ricevere e trasmettere immagini di documenti.

finestra di dialogo—In Microsoft Windows, indica una categoria dello schermo di interfaccia utente che richiede l'interazione ("dialogo") con un utente dell'applicazione. TextBridge visualizza una finestra principale dalla quale è possibile definire e avviare le operazioni di OCR.

formato bozza di stampa—Formato ad una colonna in cui TextBridge produce il testo riconosciuto tratto da documenti a più colonne.

formato di testo—Il formato dell'elaboratore testi, del foglio elettronico o altro formato di file in cui il testo riconosciuto può essere convertito ed emesso. TextBridge può produrre il testo riconosciuto in uno dei formati seguenti:

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

formato di uscita del testo—Vedere **formato di testo**.

- G** **gruppo di lingua**—Componente di TextBridge che consente all'applicazione di eseguire l'OCR su documenti composti in una lingua particolare. Selezionando l'opzione adeguata nella finestra di dialogo Preferenze di TextBridge, è possibile specificare che il documento da riconoscere è in inglese, francese, italiano, tedesco o spagnolo. TextBridge carica il gruppo della lingua adeguato prima di avviare il riconoscimento. Vedere anche **lingua di riconoscimento**.
- H** **hypertext (ipertesto)**—Funzione che permette di scorrere una base di dati di testo (un sistema di Guida in linea, ad esempio) in vari modi: selezionando un argomento in un indice; scorrendo in sequenza in avanti e all'indietro; tramite il contesto (ovvero, facendo clic su una parola contenuta in un testo per visualizzarne la definizione o un altro schermo di informazioni ad essa correlate). TextBridge utilizza il meccanismo della Guida di Microsoft Windows per percorrere il suo sistema di Guida incorporato e basato sull'hypertext.
- I** **immagine di fax**—Rappresentazione di una pagina sotto forma di dati binari (di solito con risoluzione di 200x100 o 200x200 dpi) trasmessi da una macchina facsimile (fax) o da una scheda fax modem. I computer dotati di schede fax modem sono in grado di ricevere un'immagine di fax e memorizzarla in linea sotto forma di file TIFF. TextBridge apre e riconosce il testo proveniente da un'immagine di fax in linea memorizzato in formato TIFF.
- immagine di pagina**—Immagine binaria (in bianco e nero) di una pagina registrata nella memoria del computer o sul disco. Le immagini di pagina vengono digitalizzate o lette a partire da un file TIFF e inviate a TextBridge per il riconoscimento ottico dei caratteri (OCR).

immagine TIFF—Rappresentazione binaria di una pagina o di un grafico memorizzata in formato Tagged-Image File Format (TIFF), uno standard industriale di formato dei file di immagini. TextBridge è in grado di riconoscere il testo contenuto in pagine di immagini memorizzate nelle diverse varianti TIFF che seguono:

TIFF non compresso (intestazione Intel)
TIFF CCITT-3 (intestazione Intel)
TIFF CCITT-4 (intestazione Intel)

TIFF non compresso (intestazione Motorola)
TIFF CCITT-3 (intestazione Motorola)
TIFF CCITT-4 (intestazione Motorola)

TIFF (intestazione Intel FAXability™)

Quando si seleziona **Salva immagini di pagina** nella finestra di dialogo principale, automaticamente TextBridge salva le immagini di pagina digitalizzate in file di formato TIFF CCITT-3 Intel.

interfaccia utente naturale—Nella specifica **TWAIN**, indica la serie di visualizzazioni sullo schermo e dei controlli di tastiera che un driver di sorgenti TWAIN fornisce ai programmi che supportano l'unità TWAIN. Ad esempio, TextBridge può essere eseguito con scanner dotati di driver di sorgenti totalmente conformi con lo standard TWAIN. Tuttavia, i controlli dello scanner in questione sono forniti nell'interfaccia utente naturale, non in TextBridge.

ISIS—Abbreviazione di **Image and Scanner Interface Standard** (Standard di interfaccia scanner e immagini) creato dalla Pixel Translations, Inc. ISIS è un'interfaccia di programmazione delle applicazioni (API) per l'ideazione e lo sviluppo dei driver di scanner. La Pixel Translations e altri produttori di scanner usano l'interfaccia ISIS per sviluppare i driver di scanner. TextBridge supporta la maggior parte degli scanner compatibili con ISIS attualmente disponibili sul mercato. Vedere anche **TWAIN**.

istruzione DEVICE—Nel file `config.sys`, indica una riga che identifica, ad esempio, il driver di periferica di uno scanner per le applicazioni che hanno necessità di utilizzare lo scanner.

L lingua di riconoscimento—La lingua originale (ad esempio, inglese o francese) in cui è stato composto il documento. In TextBridge si può specificare che il documento contiene testo in lingue diverse. TextBridge carica il **gruppo di lingua** adeguato prima di iniziare l'OCR.

luminosità dello scanner—Impostazione che controlla l'intensità della luce che lo scanner proietta sulla pagina sottoposta a scansione, permettendo di schiarire o scurire l'immagine risultante. Tramite la regolazione della luminosità, spesso è possibile modificare l'accuratezza di riconoscimento (ad esempio, aumentando la luminosità per una pagina che presenta caratteri poco spaziati, il riconoscimento può dare risultati migliori). In TextBridge, la luminosità dello scanner viene definita nella finestra di dialogo Impostazioni scanner.

M **maniglie**—Elementi solidi quadrati di solito presenti ai quattro angoli di un rettangolo nei pacchetti di disegno o in altre applicazioni, che consentono di modificare le dimensioni del rettangolo. In TextBridge, si può disegnare una **zona** rettangolare sull'immagine di pagina previsualizzata per definire un'area della pagina da riconoscere. Le maniglie sulla zona permettono di modificarne le dimensioni.

memoria virtuale—Spazio del disco fisso che simula la memoria ad accesso casuale (RAM) sul PC. TextBridge, soprattutto nei sistemi con solo 4 Mb di RAM, deve essere configurato con la memoria virtuale **permanente**, un gruppo contiguo di blocchi di memorizzazione contenuti nel disco fisso.

memoria virtuale permanente—Vedere **memoria virtuale**.

mezzi toni—Immagini composte da punti neri di dimensioni differenti, spaziate in modo da simulare le diverse tonalità di grigio delle fotografie o dei disegni a colori dell'originale. Selezionando un comando nella finestra di dialogo Preferenze, è possibile fare sì che TextBridge ignori i mezzi toni durante l'esecuzione del riconoscimento ottico dei caratteri.

modalità Avanzata—La più avanzata delle tre modalità in cui Microsoft Windows può essere eseguito. Le altre modalità sono Reale e Standard. TextBridge viene eseguito solo in modalità Avanzata.

N nome del documento—Il nome del file che viene immesso nella finestra di dialogo di TextBridge Salva immagine di pagina con nome o Salva con nome. Al nome del documento viene aggiunta automaticamente un'estensione di tre lettere che indica il formato in cui viene salvato il file riconosciuto.

nome di base—La porzione del **nome del documento** utilizzata per identificare i nomi dei file (TIFF) delle immagini di pagina correlati, creati tramite il comando Salva immagini di pagina di TextBridge. Quando si digita il nome di un documento nella finestra di dialogo Salva immagine di pagina con nome, le prime quattro lettere del nome vengono usate come nome di base per i file delle immagini di pagina.

O orientamento automatico di pagina—In TextBridge, indica la possibilità di correggere la rotazione dell'immagine di pagina prima dell'avvio del riconoscimento. Ad esempio, se l'utente digitalizza un documento con pagine miste (ad esempio, in maggioranza ad orientamento verticale e alcune ad orientamento orizzontale con grandi tabelle), TextBridge può applicare l'orientamento automatico ad ogni pagina, prima di iniziare il riconoscimento.

OS/2—Sistema operativo grafico ideato dalla International Business Machines Corporation (IBM®) per essere utilizzato su personal computer dotati di Intel. OS/2 è un vero sistema operativo di multielaborazione che può eseguire programmi locali e programmi ideati per Microsoft Windows e DOS.

P **parola incerta**—Parola che non rientra nella soglia di affidabilità di cui è dotato TextBridge. Durante l'OCR, TextBridge assegna ad ogni parola un valore di affidabilità. Se il valore è al di fuori della soglia di affidabilità e si sta utilizzando il **Verificatore**, TextBridge considera la parola come parola incerta.

pixel—Abbreviazione di “picture element” (elemento d'immagine); indica uno dei tanti puntini che formano l'immagine digitalizzata.

preferenze—In TextBridge, le impostazioni che possono essere specificate per controllare l'operazione OCR.

programma TSR—Programma che si carica automaticamente nella memoria quando si avvia il sistema o rimane in memoria anche quando si esce da esso.

R **regione**—Blocco logico di testo sull'immagine di pagina. Tramite la funzione di **segmentazione automatica**, TextBridge localizza regioni di testo su un documento a più colonne e le emette nell'ordine esatto.

riconoscimento—Operazione di TextBridge durante la quale un'immagine di pagina (TIFF digitalizzata o in linea) viene analizzata e i caratteri e le parole vengono identificate e salvate come flusso di dati di testo nella memoria o in un file in linea temporaneo. Tenere presente che il meccanismo di riconoscimento di TextBridge non solo esegue il riconoscimento (OCR), ma esegue anche la segmentazione, l'orientamento (rotazione), l'analisi del formato e il mantenimento degli stili del testo (grassetto, corsivo).

riconoscimento ottico dei caratteri (optical character recognition, OCR)—Tecnologia in cui le immagini binarie dei caratteri vengono analizzate e identificate come caratteri particolari ed emesse su un flusso di dati di testo, nella memoria del computer o in un file del computer.

risoluzione—Il grado di dettaglio, misurato in **punti per pollice** (dpi, dots per inch), con cui uno scanner o una macchina fax può introdurre in memoria un'immagine. TextBridge può eseguire l'OCR (riconoscimento ottico dei caratteri) sulle immagini di pagina in una delle risoluzioni seguenti (in punti per pollice): 400x400, 400x200, 300x300, 200x200, 200x100 e 100x200.

RTF—Rich Text Format, un formato di testo sviluppato dalla Microsoft Corporation, con codici annidati per la descrizione di font, la formattazione, ecc.

S segmentazione automatica di pagina—In TextBridge, indica la possibilità di esaminare lo schema dell'immagine di pagina, e di riconoscere e produrre il testo nell'ordine esatto. Ad esempio, nel caso di bollettini, in cui le colonne sono spesso di profondità e larghezze diverse, TextBridge riconosce lo schema della pagina e produce il testo nell'ordine esatto. Nelle Preferenze di TextBridge, è possibile attivare o disattivare la segmentazione automatica di pagina.

Setup scanner—Fa parte del gruppo di programmi TextBridge OCR in Windows, ed è un programma che consente di caricare il driver di alto livello adeguato in modo che TextBridge possa essere eseguito con lo scanner utilizzato. Vedere anche **ISIS** e **TWAIN**.

sorgente di entrata—L'origine delle immagini di pagina sottoposte a riconoscimento. Può essere lo scanner o un file TIFF.

stampante OCR—Applicazione TextBridge che svolge il ruolo di stampante praticamente in tutti i programmi per fax o di immagini basati su Windows. Nel programma ospite, si visualizza un fax o un'altra immagine contenente testo e si utilizza il comando Stampa per inviarla alla stampante OCR; quest'ultima esegue TextBridge OCR, poi visualizza la finestra di dialogo Salva con nome che consente di salvare in un file, nel formato desiderato, il testo riconosciuto.

T **TWAIN**—Standard di interfaccia di immagini e scanner, dotato di interfaccia API, per lo sviluppo di interfacce utilizzate dai dispositivi di elaborazione di immagini (scanner, macchine fax, ecc.). TextBridge supporta tutti gli scanner totalmente conformi allo standard TWAIN o qualsiasi dispositivo collegabile al PC che produca immagini binarie (in bianco e nero) nella risoluzione e nelle dimensioni supportate.

V **verificatore di parole**—Vedere **Verificatore**.

Verificatore—Funzione che permette di visualizzare e, se necessario, correggere parola per parola le decisioni di riconoscimento di TextBridge. La finestra del Verificatore è simile a quella di Anteprima e visualizza la parola riconosciuta e la sua immagine associata sulla pagina digitalizzata. La parola riconosciuta appare evidenziata in una casella di modifica, in cui vanno digitate le eventuali correzioni.

W **Windows**—Interfaccia utente grafica (GUI, Graphical User Interface) e una serie di moduli correlati creati dalla Microsoft Corporation per l'utilizzo su personal computer con il DOS. TextBridge può essere eseguito con Windows, versione 3.1 e successive.

Z **zona**—Nella finestra di Anteprima di TextBridge, indica un margine rettangolare che può essere disegnato attorno ad una porzione dell'immagine di pagina visualizzata, per definire l'area della pagina da elaborare.

zoom—Nella finestra di Anteprima di TextBridge, indica la possibilità di ingrandire l'immagine di pagina fino alla risoluzione massima e di ridurla fino alla risoluzione che permette di visualizzare tutta l'immagine di pagina.