



GLOSSAIRE

Ce glossaire définit des termes et concepts employés dans ce manuel. Les lecteurs qui ne possèdent pas une grande expérience de la numérisation et de la reconnaissance optique de caractères pourront l'utiliser pour se familiariser avec ces technologies.

Les termes en caractères **gras**, qui figurent dans certaines définitions, sont également définis dans ce glossaire.

A **application**—Programme logiciel qui permet aux utilisateurs d'accomplir une tâche ou un ensemble de tâches. On parle parfois de l'utilisation (c'est-à-dire de l'"application") d'un programme logiciel.

ASCII—Acronyme de American Standard Code for Information Interchange. Le jeu ASCII contient des codes pour 128 caractères de contrôle, caractères alphanumériques et symboles. Il existe également des jeux ASCII dits étendus, qui proposent généralement 128 codes additionnels pour les caractères accentués et des symboles supplémentaires.

B **boîte d'édition**—Convention de l'interface Windows ; un champ rectangulaire dans une boîte de dialogue ou dans une autre zone de l'interface, dans lequel l'utilisateur peut entrer du texte ou éditer le texte existant. Dans Windows, une boîte d'édition supporte des méthodes standard permettant à l'utilisateur d'éditer les informations qu'elle contient. Par exemple, dans la boîte de dialogue conventionnelle Enregistrer sous d'une application Windows, la zone dans laquelle le nom du fichier de sortie est indiqué est une boîte d'édition standard.

boîte de dialogue— Dans Microsoft Windows, des fenêtres de l'interface utilisateur qui apparaissent temporairement pour solliciter des informations. TextBridge affiche une boîte de dialogue principale à partir de laquelle vous pouvez définir et lancer le processus de reconnaissance optique de caractères (OCR).

C **cadre**— Dans la fenêtre Visualiser de TextBridge, un rectangle dont vous pouvez encadrer une portion de l'image de page affichée pour définir la zone à traiter.

CCITT— Acronyme de Consultative Committee of International Telephone and Telegraph, organisation internationale qui a établi, entre autres, des normes de compression des données numériques. Les fichiers TIFF stockés selon les normes de compression CCITT des groupes 3 et 4 peuvent être reconnus par TextBridge.

Configuration du scanner— Élément du groupe de programmes TextBridge OCR dans Windows. Ce programme est conçu pour vous permettre de charger le questionnaire approprié de sorte que TextBridge puisse fonctionner avec votre scanner. Voir aussi **ISIS** et **TWAIN**.

conversion— Module logiciel capable de convertir le format d'un texte. Dans TextBridge, le texte reconnu dans son format original peut être converti dans l'un des nombreux formats supportés (WordPerfect, par exemple).

couleur invisible— Couleur ou gamme de couleurs qu'un scanner a des difficultés à détecter sur la page qu'il est en train de numériser. Cette couleur dépend généralement de celle du faisceau du scanner. Par exemple, un scanner dont le faisceau est jaune détectera difficilement les couleurs allant du jaune au vert clair.

D DDE— Voir échange de données dynamique.

demi-ton—Image composée de points noirs de tailles diverses, espacés de façon à simuler les différents tons de gris d'un dessin en couleurs ou d'une photographie originale. La boîte de dialogue Préférences comporte une option qui permet de contraindre TextBridge à ignorer les demi-tons au cours de la reconnaissance optique de caractères.

disque RAM—Partie de la mémoire étendue de votre ordinateur configurée afin de fonctionner comme un disque dur pour le stockage de fichiers temporaires.

E échange de données dynamique (DDE)—Dans l'environnement Microsoft Windows, une norme de partage des données entre applications Windows. Par exemple, avec une macro appropriée, un programme de traitement de texte (tel que Microsoft Word pour Windows) peut demander à TextBridge de numériser et de reconnaître un texte, et de l'importer dans un document Word ouvert, à partir de son propre système de menus.

F fax modem—Périphérique ou carte de circuits imprimés qui se connecte à un PC pour permettre la réception et la transmission de données numériques via une ligne téléphonique. Avec un fax modem reliant votre PC à une ligne téléphonique, vous pouvez recevoir/envoyer des images de documents sur/à partir de votre PC.

format texte—Format de traitement de texte, de tableur ou autre format de fichier dans lequel le texte reconnu peut être converti et généré. TextBridge supporte les formats suivants :

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 de Microsoft)
DisplayWrite 5	Samna Word IV
Excel (Mac, 3.0, 4.0)	Windows Write
FrameMaker	Word pour Windows 2.0
Interleaf	WordPerfect (4.2, 5.1)
Lotus 1-2-3	WordStar

G galée, format—Format à une seule colonne dans lequel TextBridge génère le texte reconnu d'un document qui comporte plusieurs colonnes.

gestionnaire de mémoire commutée—Programme qui convertit une partie de la mémoire étendue en mémoire commutée additionnelle, de sorte que les programme qui requièrent plus que les 360 Ko de mémoire commutée, généralement disponibles sur les machines DOS puissent être exécutés.

gestionnaire de scanner—Programme écrit en tant qu'interface entre une application logicielle et un scanner. Le gestionnaire de scanner envoie les requêtes de l'application au scanner dans un langage que ce dernier peut comprendre. (On parle aussi de driver.)

H **hypertext**—Fonctionnalité permettant d'explorer une base de données textuelle de différentes façons (un système d'aide en ligne par exemple) : par sélection d'un sujet dans un index ; par déplacements séquentiels vers l'avant ou l'arrière ; par recherche de mot-clé ; par contexte (c'est-à-dire en cliquant sur un mot dans le texte pour afficher sa définition ou un autre écran d'information). TextBridge utilise l'Aide de Microsoft Windows pour parcourir son système d'aide intégré, basé sur hypertext.

I **image de fax**—Représentation d'une page sous la forme de données binaires (en général, dans une résolution de 200x100 ou 200x200 ppp) transmises par un télécopieur ou une carte de fax modem. Les ordinateurs équipés d'une telle carte peuvent recevoir une image de fax et la stocker en ligne en tant que fichier TIFF. TextBridge peut ouvrir et reconnaître le texte d'une image de fax en ligne, stockée en format TIFF.

image de page—Image binaire (noir et blanc) d'une page stockée en mémoire ou sur disque. Les images de pages sont numérisées ou lues à partir d'un fichier TIFF puis envoyées à TextBridge pour la reconnaissance optique de caractères (OCR).

image TIFF—Représentation binaire d’une page ou d’un graphique stocké en format “Tagged-Image File Format”, un format de fichier d’image standard. TextBridge peut reconnaître le texte des images stockées dans les versions suivantes du format TIFF :

TIFF non compressé (en-tête Intel)
TIFF CCITT-3 (en-tête Intel)
TIFF CCITT-4 (en-tête Intel)

TIFF non compressé (en-tête Motorola)
TIFF CCITT-3 (en-tête Motorola)
TIFF CCITT-4 (en-tête Motorola)

TIFF (en-tête Intel FAXability™)

Lorsque vous sélectionnez Enregistrer images de pages dans la boîte de dialogue principale, TextBridge enregistre automatiquement les images numérisées dans des fichiers TIFF CCITT-3 Intel.

Imprimante OCR—Application de TextBridge conçue pour fonctionner comme une imprimante à partir de virtuellement tout programme de télécopie ou d’imagerie articulé autour de Windows. Dans le programme hôte, affichez un fax ou une autre image contenant un texte puis utilisez la commande Imprimer pour l’envoyer à l’Imprimante OCR. Cette dernière exécute le programme de reconnaissance optique de TextBridge, puis affiche la boîte de dialogue Enregistrer sous pour vous permettre d’enregistrer le texte dans un fichier au format de votre choix.

instruction DEVICE—Dans le fichier `config.sys`, une ligne qui identifie, par exemple, un gestionnaire de scanner pour une application susceptible d’utiliser le scanner.

interface utilisateur originale— Dans les spécifications **TWAIN**, les différents écrans et commandes de contrôle du clavier qu'un gestionnaire source TWAIN fournit aux programmes supportant une unité TWAIN. Par exemple, TextBridge fonctionne avec les scanners dont les gestionnaires source sont totalement conformes à la norme TWAIN. Cependant, les commandes de contrôle d'un tel scanner sont fournies par son interface utilisateur originale, et non par TextBridge.

ISIS— Acronyme de **I**mage and **S**canner **I**nterface **S**tandard, norme développée par Pixel Translations, Inc. ISIS est une interface API (Applications Programmers Interface) destinée à la conception et au développement des gestionnaires de scanners. Pixel Translations et de nombreux autres fabricants de scanners utilisent ISIS pour développer des gestionnaires. TextBridge supporte la plupart des scanners compatibles ISIS disponibles sur le marché. Voir aussi **TWAIN**.

L **langue de reconnaissance**— Langue principale (par exemple, l'anglais ou le français) dans laquelle un document est écrit. Choisissez cette langue dans TextBridge : il chargera le **module de langue** approprié avant de commencer la reconnaissance.

luminosité— Paramètre déterminant l'intensité de la lumière que le scanner projette sur la page numérisée afin d'éclaircir ou d'assombrir l'image résultante. En ajustant la luminosité, vous pouvez améliorer la précision de la reconnaissance (par exemple, en éclaircissant une page dont les caractères sont très serrés). Vous avez la possibilité de spécifier la luminosité du scanner dans la boîte de dialogue Paramètres du scanner de TextBridge.

M **mémoire virtuelle**—Espace de votre disque dur configuré pour simuler la mémoire RAM de votre PC. TextBridge, en particulier sur les systèmes ne disposant que de 4 Mo de mémoire RAM, doit être configuré avec une mémoire virtuelle **permanente**, un groupe contigu de blocs de stockage sur votre disque dur.

mode Etendu—Le plus avancé des trois modes dans lesquels Microsoft Windows peut être exploité. Les deux autres modes sont Réel et Standard. TextBridge fonctionne uniquement en mode Etendu.

module de langue—Composante de TextBridge qui permet à l'application d'effectuer la reconnaissance optique des caractères d'un document écrit dans une langue particulière. Vous pouvez choisir l'une des langues suivantes dans la boîte de dialogue Préférences de TextBridge: Anglais, Français, Allemand, Italien ou Espagnol. TextBridge charge le module de langue approprié avant de procéder à la reconnaissance. Voir aussi **langue de reconnaissance**.

mot suspect—Mot en dessous du seuil de fiabilité défini dans TextBridge. Au cours de la reconnaissance optique, TextBridge assigne une valeur de fiabilité à chaque mot. Si cette valeur est inférieure au seuil de fiabilité et que vous utilisez le **Vérificateur**, TextBridge considère le mot correspondant comme suspect et l'affiche en tant que tel.

N **nom de base**—La partie du **nom du document** utilisée pour identifier des noms de fichiers TIFF ayant un caractère commun, créés à l'aide de la commande Enregistrer images de pages de TextBridge. Lorsque vous tapez un nom de document dans la boîte de dialogue Enregistrer image de page sous, les quatre premiers caractères constituent le nom de base des fichiers d'images de pages.

nom du document—Nom de fichier que vous entrez dans la boîte de dialogue Enregistrer image de page sous ou Enregistrer sous de TextBridge. Une extension de trois caractères est automatiquement ajoutée à ce nom pour indiquer le format dans lequel le texte reconnu est enregistré.

O orientation automatique—Dans TextBridge, fonctionnalité qui permet de corriger l’orientation d’une image de page avant sa reconnaissance. Par exemple, si un utilisateur numérise un document comportant des pages “portrait” et des pages “paysage”, TextBridge peut orienter chaque page automatiquement avant de procéder à la reconnaissance.

OS/2—Système d’exploitation graphique conçu par International Business Machines Corporation (IBM®) pour fonctionner sur les PC équipés d’un processeur Intel. OS/2 est un véritable système d’exploitation multitraitement pouvant exécuter les programmes originaux aussi bien que ceux écrits pour Microsoft Windows et DOS.

P parasite—Marque sur une page, qui risque d’être reconnue en tant que caractère(s) au cours de la reconnaissance optique.

pixel—Abréviation de l’expression “picture element” (élément d’image). L’un des points composant une image numérique.

préférences—Dans TextBridge, paramètres que vous spécifiez pour contrôler le processus de reconnaissance optique.

Programme résident (TSR)—Programme conçu pour se charger automatiquement en mémoire lorsque vous démarrez votre système, ou pour “résider” en mémoire après que vous l’avez quitté.

R **reconnaissance**—Processus au cours duquel une image de page (numérisée ou TIFF en ligne) est analysée ; les caractères et les mots qu'elle contient sont identifiés et enregistrés en tant que données texte, en mémoire ou dans un fichier temporaire en ligne. Notez qu'en plus de la reconnaissance (OCR), TextBridge exécute les diverses tâches liées à la segmentation, à l'orientation, à l'analyse du format et au maintien des attributs du texte (gras, italique).

reconnaissance optique de caractères (OCR)—Technologie qui permet d'analyser et d'identifier des images binaires en tant que caractères particuliers, et de les extraire sous forme de texte, en mémoire ou dans un fichier.

région—Bloc logique de caractères sur une image de page. TextBridge, grâce à la fonctionnalité de **segmentation de page automatique**, identifie les régions de texte d'un document qui comporte plusieurs colonnes et les génère dans l'ordre approprié.

repères—Carrés noirs, affichés aux quatre coins d'un cadre, dans un progiciel de dessin ou une autre application, qui permettent de redimensionner le cadre. Dans TextBridge, vous pouvez tracer un **cadre** sur une image de page pour définir la zone à reconnaître. Les repères vous permettent de redimensionner cette zone.

répertoire de travail—Dans Windows, lors de l'installation d'une application ou à n'importe quel moment par la suite, vous pouvez désigner un répertoire de votre système de fichiers DOS comme répertoire de travail. Cette opération s'effectue dans le Gestionnaire de programmes en sélectionnant la commande Propriétés du menu Fichier. Pour TextBridge, le répertoire de travail est le sous-répertoire BIN du répertoire d'installation (spécifié au moment de l'installation). Le programme d'installation de TextBridge choisit par défaut C:\TXBRIDGE\BIN. Vous pouvez sélectionner n'importe quel autre répertoire de votre système de fichiers DOS.

résolution—Degré de détail, mesuré en **points par pouce** (ppp), avec lequel un scanner ou un télécopieur peut traiter une image. TextBridge effectue la reconnaissance optique de caractères sur les images de pages qui lui sont transmises dans les résolutions suivantes (en points par pouce) : 400x400, 400x200, 300x300, 200x200, 200x100 et 100x200.

RTF—Rich Text Format, un format texte développé par Microsoft Corporation, qui intègre des codes décrivant les polices, le formatage etc.

S **segmentation de page automatique**—Dans TextBridge, fonctionnalité qui permet de discerner la mise en page d'une image, puis de reconnaître et de générer le texte qu'elle contient dans le bon ordre. Par exemple, dans le cas d'un document imprimé comportant des colonnes de diverses tailles, TextBridge reconnaît la mise en page et génère le texte dans l'ordre approprié. Vous pouvez activer ou désactiver cette fonctionnalité dans la boîte de dialogue Préférences de TextBridge.

source—Origine des images de pages reconnues : scanner ou fichier TIFF.

- T** **TWAIN**—Norme incluant une API (Applications Programmers Interface) destinée au développement d’interfaces de périphériques d’imagerie (scanners, télécopieurs, etc.). TextBridge est compatible avec tout scanner ou autre unité totalement conforme à la norme TWAIN, qui se connecte à un PC et génère des images binaires (noir et blanc) dont la taille et la résolution sont supportées.
- V** **Vérificateur**—Fonctionnalité qui permet de visualiser et, si nécessaire, de corriger les décisions de reconnaissance de TextBridge, mot par mot. La fenêtre Vérificateur, similaire à la fenêtre Visualiser, affiche le mot reconnu dans une boîte d’édition, pour que vous puissiez le corriger si besoin est. Elle l’affiche également en surbrillance dans son contexte.
- Visualiser**—Dans TextBridge, fonctionnalité qui permet de visualiser, d’agrandir/réduire et d’encadrer des portions de pages avant la reconnaissance optique.
- W** **Windows**—Interface utilisateur graphique (GUI) et hôte de modules développés par Microsoft Corporation pour une utilisation sur les PC articulés autour du DOS. TextBridge fonctionne avec la version 3.1 (et les versions ultérieures) de Windows.
- Z** **zoom**—Dans la fenêtre Visualiser de TextBridge, une fonctionnalité qui permet d’agrandir une page (“zoom avant”) jusqu’à sa résolution maximale ou de la réduire (“zoom arrière”) de sorte que son image entière puisse être visualisée.