3 ARBEITEN MIT TEXTBRIDGE

Nach der Installation von TextBridge, wie sie in Kapitel 2 beschrieben wird, können Sie die Anwendung nun dazu benutzen, Papierdokumente und Online-Seitenbilder in verwendbare Textdateien umzuwandeln.

TextBridge kann eine Vorlage scannen, Online-**TIFF-**Grafiken lesen und eine **Optische Zeichenerkennung** (OCR) durchführen.

Die Seiten können vor der Verarbeitung in einem **Vorschaufenster vergrößert**, **verkleinert** oder in **Ausschnitten angezeigt** werden.

Während der OCR können Sie erkannte Wörter **prüfen**, korrigieren und Erkennungsfehler beheben. Dabei bringen Sie TextBridge gleichzeitig bei, die Genauigkeit im weiteren zu verbessern.

Nach der OCR können Sie mit Hilfe der Anwendung den Text in ein **Textformat** konvertieren, das Sie in verschiedenen Anwendungen (Textverarbeitung, Desktop-Publishing, Tabellenkalkulation, Datenbanken, etc.) verwenden können.

Dieses Kapitel beschreibt Schritt für Schritt, wie man mit der Haupt-TextBridge-Anwendung arbeitet.

+ Informationen zum TextBridge-Anwendungs-Server und OCR Printer finden Sie in Kapitel 4.

Dieses Kapitel beschäftigt sich mit den folgenden Themen:

- TextBridge-Voreinstellungen
- Scannen und Konvertieren einer Vorlage
- Erkennen und Konvertieren von Online-TIFF-Dateien
- Seitenvorschau vor der Verarbeitung
- Textprüfung während der Erkennung

TEXTBRIDGE-VOREINSTELLUNGEN

Die TextBridge-Software zur optischen Zeichenerkennung verfügt über integrierte **Voreinstellungen**. Dabei handelt es sich um Einstellungen, die den OCR-Vorgang steuern. Für die meisten hochwertigen Bürodokumente reichen die Standardvoreinstellungen von TextBridge völlig aus.

Mit TextBridge können Sie den OCR-Vorgang aber auch auf einzelne Vorlagen genau abstimmen. Klicken Sie im Hauptdialogfeld auf die Schaltfläche "Voreinstellungen", um das entsprechende Dialogfeld anzuzeigen (Abbildung 3–1).

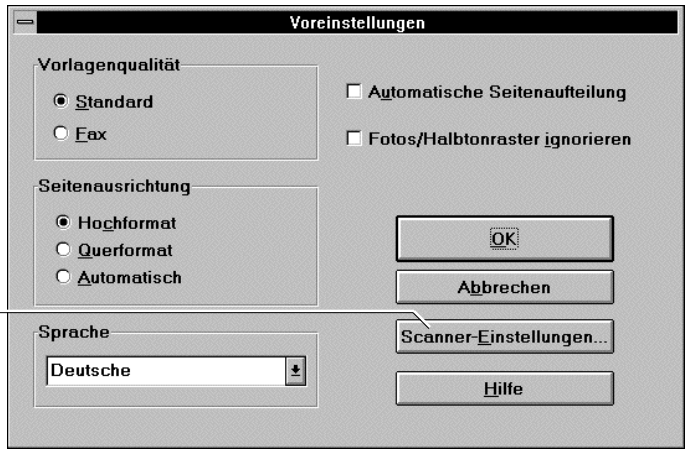

Abbildung 3–1. Das Dialogfeld "Voreinstellungen"

Bestimmen Sie nach Ihren Wünschen eine oder mehr der Voreinstellungen. Tabelle 3–1 beschreibt die Ihnen verfügbaren Optionen.

Hinweis Eine von Ihnen veränderte Voreinstellung wird so lange beibehalten, bis Sie sie erneut ändern, auch wenn Sie TextBridge verlassen und das Programm später wieder verwenden. Auf diese Weise können Sie die für Sie wesentlichen Voreinstellungen auf längere Zeit festlegen. Falls notwendig, können sie natürlich jederzeit wieder geändert werden.

Tabelle 3–1. Voreinstellungen

Voreinstellung	Funktion
Vorlagenqualität	• Wählen Sie **Standard** für die meisten Vorlagen. • Wählen Sie **Fax** für Online-TIFF-Grafiken (200x100 oder 200x200 dpi) oder beim Scannen von Fax-Papiervorlagen.
Seitenausrichtung	• Klicken Sie bei typischen hochformatigen Bürodokumenten auf **Hochformat**. • Klicken Sie bei breiten, quer zu scannenden Vorlagen, die natürlich vor der OCR um 90 Grad rotiert werden müssen, auf **Querformat**. • Klicken Sie auf **Automatisch**, wenn Ihre Vorlage eine Mischung von Seitenausrichtungen enthält (z.B. viele Hochformate, aber einige große, rotierte Tabellen auf den querformatigen Seiten). Klicken Sie auch dann auf “Automatisch”, wenn Sie bei TIFF-Dateien nicht sicher sind, wie das Seitenbild in der Datei ausgerichtet ist. Wenn “Automatisch” ausgewählt ist, führt TextBridge eine Vorverarbeitung durch, bei der die Seitenausrichtung festgestellt wird. Die Verarbeitungsgeschwindigkeit wird daher verlangsamt, wenn diese Option aktiviert ist.

Tabelle 3–1. Voreinstellungen (Forts.)

Voreinstel-lung	Funktion
Sprache	Ein Popup-Menü listet die Sprachen auf, die Sie auf Ihrem System installiert haben. TextBridge führt eine hochwertige optische Zeichenerkennung bei deutschen, französischen, italienischen und spanischen Vorlagen durch. Wählen Sie die primär zu erkennende Sprache aus (z.B. deutsch).
Automatische Seiten-aufteilung	• Bei einspaltigen Vorlagen **deaktivieren**. • Bei zwei- oder mehrspaltigen Vorlagen **aktivieren**. Wenn diese Option eingeschaltet ist, führt TextBridge eine Vorverarbeitung zur Analyse von Form und Position der Textblöcke auf der Seite durch, damit diese in der richtigen Reihenfolge ausgegeben werden. In der konvertierten Textdatei gibt TextBridge Text im **Satzschiff-** (d.h. einspaltigen) Format aus.
Fotos/ Halbton-raster ignorieren	Klicken Sie auf das Kontrollkästchen **Fotos/Halbtonraster ignoroieren**, wenn das zu erkennende Dokument Fotos und Halbtonraster enthält. Andernfalls versucht TextBridge, die Halbtonpunkte als Text zu erkennen, wodurch Erkennungsgenauigkeit und -geschwindigkeit reduziert werden.

Tabelle 3–1. Voreinstellungen (Forts.)

Voreinstel-lung	Funktion
Scanner-Einstellungen	Klicken Sie auf die Schaltfläche “Scanner-Einstellungen”, um das entsprechende Dialogfeld anzuzeigen. • Die Option **Automatische Vorlagenzuführung** ist nur bei den entsprechenden Scannern verfügbar. Wenn Sie auf das Kontrollkästchen klicken, werden die Seiten in der ADF erkannt und gescannt. • Mit **Helligkeit** können Sie die beim Scannen verwendete Lichtmenge und damit die Helligkeit der gescannten Seite bestimmen. Dadurch können Sie die unterschiedlichen Hellig-keitsstufen verschiedener Vorlagen kompensieren und den Erkennungsprozeß verbessern. + Wenn Sie einen HP-Scanner mit HP AccuPage verwenden, ist “Automatisch” die richtige Einstellung. • Mit **Auflösung** können Sie die Anzahl der Punkte pro Zoll (dpi) bestimmen, gemäß derer die Seite(n) gescannt werden soll(en). Für optimale Ergebnisse sollten Sie die höchste Auflösung wählen, die Ihr Scanner zuläßt (bis zu 400 dpi). • Mit **Seitengröße** können Sie die Größe des zu scannenden Bereichs festlegen. Geben Sie dabei möglichst die genaue Größe der zu scannenden Seite an: **US-Letter** (8,5 mal 11 Zoll oder 21,59 mal 27,94 Zentimeter); **US-Legal** (8,5 mal 14 Zoll oder 21,59 mal 35,56 Zentimeter); oder **A4** (8,27 mal 11,69 Zoll oder 21 mal 29,70 Zentimeter)

SCANNEN UND KONVERTIEREN EINER VORLAGE

Mit TextBridge können Sie u.a. eine Papiervorlage in eine Online-Textdatei umwandeln. Die Vorlage kann eine oder mehrere Seiten umfassen.

TextBridge scannt dabei die erste Seite, führt eine OCR durch und speichert den erkannten Text in einer temporären Datei. Dieser Vorgang wird für jede Seite wiederholt, und der jeweils neue Text an die temporäre Datei angehängt.

Wenn Sie TextBridge mitteilen, daß die gesamte Vorlage gescannt wurde, wird das Dialogfeld "Speichern unter" angezeigt. Hier geben Sie den Namen, das Laufwerk und das Verzeichnis sowie das Textformat der Ausgabedatei an.

Hinweis Im folgenden wird davon ausgegangen, daß der Scanner ordnungsgemäß an Ihren PC angeschlossen und eingeschaltet ist, und daß weiterhin TextBridge installiert und gestartet ist.

So scannen und konvertieren Sie eine Vorlage:

1. **Legen Sie die Seite(n) in Ihren Scanner ein.**

   Je nach Scanner können Sie einige Seiten in die automatische Vorlagenzuführung (ADF) oder eine einzelne Seite in das Flachbett des Scanners legen.

2. **Im Hauptdialogfeld klicken Sie auf Scanner in dem Feld "Eingabe aus" (siehe Abbildung 3–2 auf der nächsten Seite).**

3. **Um ein Bild jeder Seite in einer TIFF-Datei zu speichern, klicken Sie auf das Kontrollkästchen “Seite als Bild speichern”.**

   + Seitenbilder werden als TIFF-Dateien mit **CCITT**-Komprimierung (Gruppe 3) gespeichert. Weitere Informationen über die Option “Seite als Bild speichern” finden Sie in Kapitel 4.

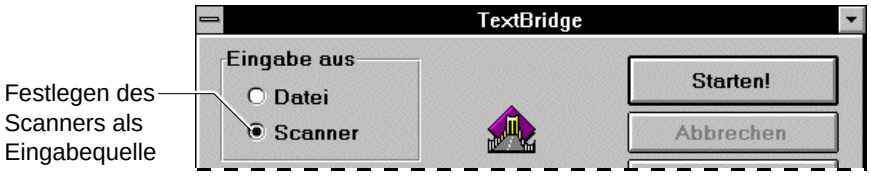

Abbildung 3–2. Scanner als Eingabequelle

4. **Wahlweise: Klicken Sie auf die Schaltfläche “Voreinstellungen”, um das Scannen und die OCR genauer zu steuern.**

   Richten Sie sich dabei nach Tabelle 3–1. Mit den Voreinstellungen können Sie z.B. die Seitenausrichtung und die zu erkennende Sprache bestimmen. Klicken Sie danach auf “OK”, um zum Hauptdialogfeld zurückzukehren.

5. **Klicken Sie dort auf “Starten!”, um mit dem Scannen und der OCR zu beginnen.**

   Wenn Sie einen TWAIN-Scanner verwenden, wird die **native Benutzerschnittstelle** des Quelltreibers angezeigt. Hier können Sie den Scan-Vorgang steuern.

   Wenn Sie einen ISIS-Scanner verwenden, werden die Seiten im Scanner automatisch verarbeitet.

Wenn Sie die native Benutzerschnittstelle schließen (TWAIN) oder die Seiten schon alle gescannt sind (ISIS), zeigt TextBridge das Dialogfeld “Weitere Seiten hinzufügen” an (Abbildung 3–3).

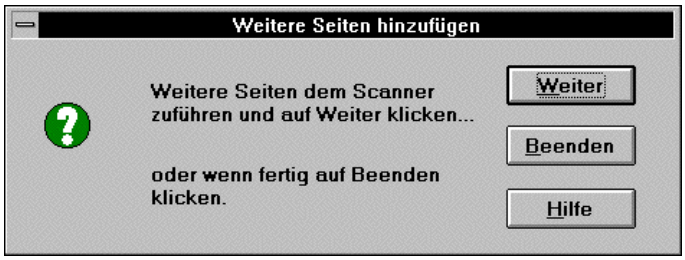

Abbildung 3–3. Das Dialogfeld “Weitere Seiten hinzufügen”

6. **Um weitere Seiten zu scannen, fahren Sie mit Schritt 7 fort. Um Scannen und OCR zu beenden, gehen Sie direkt zu Schritt 8 über.**

7. **Bereiten Sie Ihren Scanner vor, und klicken Sie auf “Weiter”.**

   Wenn Sie einen TWAIN-Scanner verwenden, wird die native Benutzerschnittstelle des Quelltreibers erneut angezeigt, und Sie können mit dem Scannen und der OCR fortfahren. Wenn Sie einen ISIS-Scanner verwenden, nimmt TextBridge automatisch die Verarbeitung wieder auf. Wenn Sie fertig sind, zeigt TextBridge das Dialogfeld “Weitere Seiten hinzufügen” an (siehe Abbildung 3–3). Beginnen Sie wieder bei Schritt 6.

8. **Klicken Sie im Dialogfeld “Weitere Seiten hinzufügen” auf “Beenden”.**

   TextBridge zeigt nun das Dialogfeld “Speichern unter” an (Abbildung 3–4).

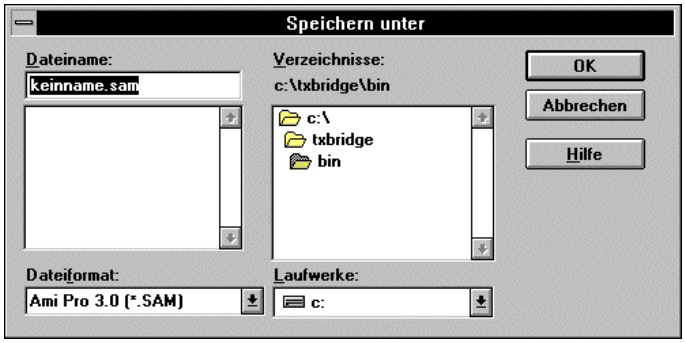

Abbildung 3–4. Das Dialogfeld “Speichern unter”

9. **Geben Sie die Ausgabeinformationen an.**

   - Bestimmen Sie den Dateinamen, das Ziellaufwerk und das Zielverzeichnis.
   - Bestimmen Sie auch das Ausgabeformat in dem Popup-Menü “Dateiformat”.
   - Klicken Sie auf “OK”.

   TextBridge konvertiert den erkannten Text in das angegebene Format, speichert ihn in der angegebenen Datei und schließt das Dialogfeld. Das Hauptdialogfeld bleibt geöffnet und ist bereit für die nächste Aufgabe.

ERKENNEN UND KONVERTIEREN VON TIFF-DATEIEN

Wenn eine oder mehrere Ihrer TIFF-Dateien Seitenbilder enthalten, können Sie mit TextBridge eine Textdatei daraus erstellen. TextBridge verarbeitet eine Online-TIFF-Grafikdatei ähnlich wie ein oder mehrere gescannte Seitenbild(er).

Manche TIFF-Dateien enthalten mehr als ein Seitenbild. Diese mehrseitigen TIFF-Dateien werden im Normalfall von **Faxmodems** generiert. TextBridge kann sowohl ein- als auch mehrseitige TIFF-Dateien verarbeiten.

TextBridge kann TIFF-Dateien (`*.TIF`) in den in Tabelle 3–2 aufgeführten Auflösungen und Formaten verarbeiten:

Tabelle 3–2. Unterstützte TIFF-Auflösungen und Formate

Auflösungen*	TIFF-Formate
100 mal 200 200 mal 100 200 mal 200 300 mal 300 400 mal 200 400 mal 400	Nicht komprimiert (Intel-Kennsatz) CCITT-3 (Intel-Kennsatz) CCITT-4 (Intel-Kennsatz) Nicht komprimiert (Motorola-Kennsatz) CCITT-3 (Motorola-Kennsatz) CCITT-4 (Motorola-Kennsatz) Intel FAXability

* Alle als eine Vorlage verarbeiteten TIFF-Dateien müssen die gleiche Auflösung aufweisen. Wenn TextBridge mehrere Dateien mit 200 mal 200 dpi verarbeitet und bei der nächsten zu verarbeitenden Datei 300 mal 300 dpi vorliegen, generiert das Programm eine Fehlermeldung und zeigt das Dialogfeld “Speichern unter” an, damit Sie die bis zu diesem Zeitpunkt erzielten Erkennungsergebnisse in einem Ausgabedokument speichern können.

Ihren Angaben folgend, öffnet TextBridge die TIFF-Datei(en), liest die Bilddaten in den Arbeitsspeicher des Computers, führt bei dem Bild/den Bildern eine OCR durch und speichert den erkannten Text in einer formatierten Textdatei.

Hinweis Im folgenden wird davon ausgegangen, daß TextBridge ordnungsgemäß installiert ist und Ihr PC eingeschaltet ist.

Um eine oder mehrere Online-TIFF-Dateien in eine Textdatei zu konvertieren, gehen Sie folgendermaßen vor:

1. **Klicken Sie im Hauptdialogfeld in dem Feld "Eingabe aus" auf "Datei" (Abbildung 3–5).**

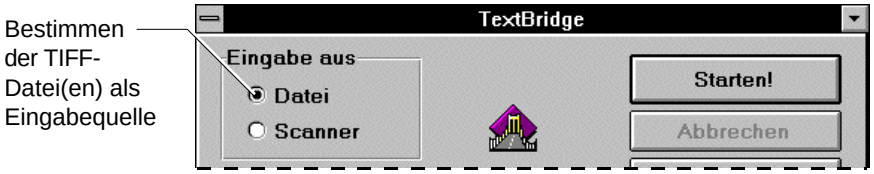

Abbildung 3–5. Eingabe einer Datei

2. **Wahlweise: Bestimmen Sie die Voreinstellungen zur Steuerung des Erkennungsvorgangs.**
   - Klicken Sie auf die Schaltfläche "Voreinstellungen", um das gleichnamige Dialogfeld anzuzeigen (siehe Abbildung 3–1).
   - Verändern Sie eine oder mehrere der Voreinstellungen. Tabelle 3–1 enthält die dafür notwendigen Informationen.
   - Wenn Sie mit dem Bestimmen von Voreinstellungen fertig sind, klicken Sie auf "OK", um zum Hauptdialogfeld zurückzukehren.

3. **Klicken Sie im Hauptdialogfeld auf "Starten!".**

   TextBridge zeigt das Dialogfeld "Öffnen" (Abbildung 3–6).

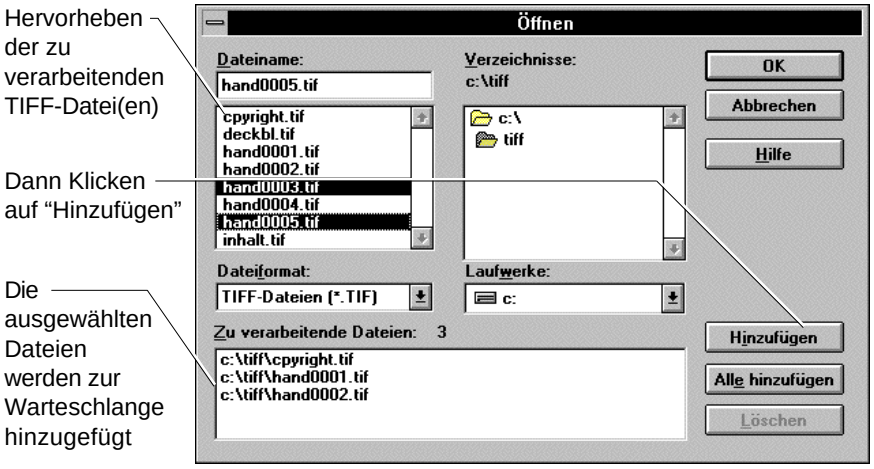

Abbildung 3–6. Das Dialogfeld "Öffnen"

4. **Geben Sie die TIFF-Datei(en) an, und klicken Sie auf "OK".**

   + Klicken Sie zur Auswahl einer einzelnen Datei auf den entsprechenden Dateinamen in der Liste, und klicken Sie auf "Hinzufügen", um diese Datei der Liste "**Zu verarbeitende Dateien**" hinzuzufügen. Oder klicken Sie doppelt auf die Datei.

     Um mehrere Dateien auszuwählen, klicken Sie auf die erste Datei, und ziehen Sie die Maus bei gedrückter Maustaste in die Dateiliste. Oder klicken Sie auf die erste Datei und klicken Sie danach bei gedrückter Umschalttaste auf die letzte Datei. Nach Auswahl der gewünschten Dateien klicken Sie auf "Hinzufügen", um die Dateien in der Liste zu verzeichnen.

+ Um nicht aufeinanderfolgende TIFF-Dateien auszuwählen, klicken Sie auf die erste Datei und anschließend bei gedrückter Steuerungstaste auf die restlichen Dateien.

Um alle Dateien im aktuellen Verzeichnis der Verarbeitungsliste hinzuzufügen, klicken Sie auf die Schaltfläche “Alle hinzufügen”.

Um Dateien aus einem anderen Verzeichnis hinzuzufügen, wechseln Sie in dieses Verzeichnis, und fügen Sie die Dateien hinzu, wie oben beschrieben. Der Liste werden dann die vollständigen Pfadbezeichnungen der Dateien hinzugefügt.

Um Dateien aus der Liste zu entfernen, wählen Sie eine nach der anderen und klicken auf die Schaltfläche “Löschen”. Sie können auch mehrere Dateien gleichzeitig löschen – gehen Sie bei der Auswahl wie oben beschrieben vor.

Die Dateien werden in der Reihenfolge ihres Hinzufügens verarbeitet. Die Anzahl der Dateien in der Liste wird von einem Zähler registriert (Abbildung 3–7).

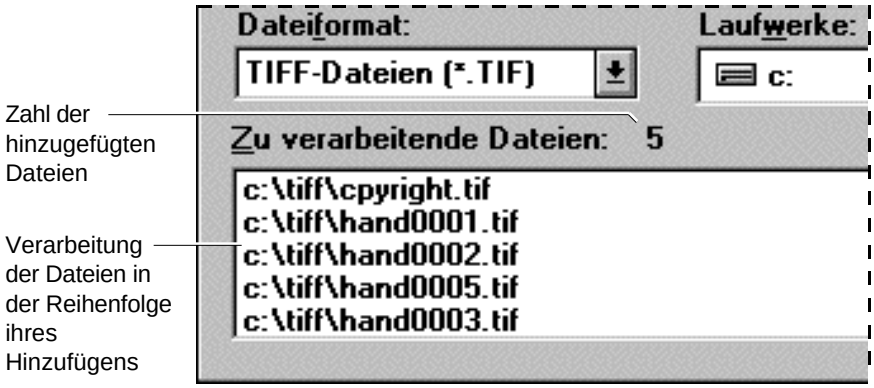

Abbildung 3–7. Mehrere zu verarbeitende TIFF-Dateien

Nachdem Sie Dateien hinzugefügt und im Dialogfeld “Öffnen” auf “OK” geklickt haben, führt TextBridge eine OCR der Seitenbilder durch. Im Anschluß daran zeigt TextBridge das Dialogfeld “Speichern unter” an (siehe Abbildung 3–4).

5. **Geben Sie im Dialogfeld “Speichern unter” Informationen zur Ausgabedatei ein.**

   - Bestimmen Sie den Dateinamen, das Ziellaufwerk und -verzeichnis.
   - Bestimmen Sie auch das Ausgabeformat in dem Popup-Menü “Dateiformat”.
   - Klicken Sie auf “OK”.

   TextBridge konvertiert den erkannten Text und speichert ihn in der angegebenen Datei im angegebenen Laufwerk und Verzeichnis. Danach wird wieder das Hauptdialogfeld angezeigt.

SEITENVORSCHAU VOR DER VERARBEITUNG

Um eine Seite vor der Verarbeitung anzuzeigen oder Bereiche einer zu verarbeitenden Seite zu bestimmen, verwenden Sie das Vorschaufenster von TextBridge (Abbildung 3–8).

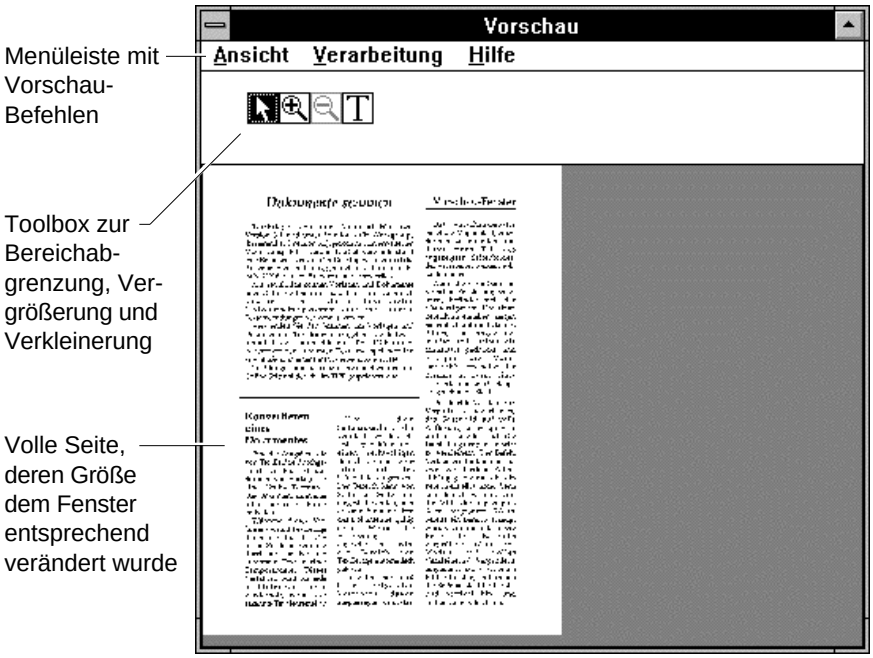

Abbildung 3–8. Das Vorschaufenster

Im Vorschaufenster können Sie die Werkzeuge zur Vergrößerung und Verkleinerung in der Toolbox dazu verwenden, die Seite stufenweise vergrößert oder verkleinert anzuzeigen:

Vergößern

Verkleinern

Mit diesen Werkzeugen können Sie eine Seite zu voller **Auflösung** vergrößern, so stark verkleinern, daß die gesamte Seite ins Fenster paßt oder eine mittlere Einstellung wählen.

Um einen zu verarbeitenden Bereich einer Seite zu bestimmen, können Sie mit dem Textbereich-Tool bis zu 127 rechteckigen **Bereichen** auf der Seite einzeichnen:

Textbereich

Die Bereiche sind entsprechend der Reihenfolge, in der sie definiert wurden (bzw. der Reihenfolge der Ausgabe), numeriert. Überlappungsbereiche sind undurchsichtig; zu dem obersten Bereich “gehört” jeweils der mehreren Bereichen gemeinsame Text.

Das Werkzeug zur Bereichsauswahl erlaubt Ihnen irgendeinen der erzeugten Bereiche auszuwählen:

Bereichsauswahl

Ein ausgewählter Bereich ist an den **Ziehpunkten** zu erkennen. Sie können einen ausgewählten Bereich verschieben, vergrößern/verkleinern, oder vor bzw. hinter einen anderen Bereich schieben.

Zur Verarbeitung können Sie für jede Seite unterschiedliche Bereiche definieren, oder Sie verwenden für alle Seiten der Vorlage die gleiche Bereichsdefinition. Nach der Verarbeitung werden die Bereiche automatisch gelöscht.

Hinweis Im folgenden wird davon ausgegangen, daß Ihr Scanner (falls Sie einen verwenden) ordnungsgemäß an Ihren PC angeschlossen und eingeschaltet ist und daß weiterhin TextBridge ordnungsgemäß installiert und gestartet ist.

So zeigen Sie ein Dokument vor der Verarbeitung an:

1. **Wenn Sie scannen wollen, legen Sie die Seite(n) in Ihren Scannerein, und gehen Sie zu Schritt 2 weiter. Wenn Sie eine TIFF-Datei verarbeiten, beginnen Sie bei Schritt 2.**

2. **Definieren Sie die Eingabequelle im Hauptdialogfeld.**

   - Wählen Sie entweder “Datei” oder “Scanner” im Feld “Eingabe aus” (siehe Abbildungs 3–2 und 3–5).
   - Klicken Sie auf das “Vorschau”-Feld im Hauptdialogfeld.

3. **Wahlweise: Klicken Sie auf die Schaltfläche “Voreinstellungen”, um das Scannen und die OCR genauer zu bestimmen.**

   Richten Sie sich dabei nach Tabelle 3–1. Wenn Sie fertig sind, klicken Sie auf “OK”, um zum Hauptdialogfeld zurückzukehren.

4. **Klicken Sie im Hauptdialogfeld auf "Starten!".**

   Wenn Sie einen TWAIN-Scanner verwenden, wird die native Benutzerschnittstelle angezeigt, von der aus Sie das Scannen durchführen können. Beim ISIS-Scanner beginnt TextBridge automatisch mit dem Scannen der Seite.

   Wenn Sie die Seitenbilder von einer oder mehreren TIFF-Dateien lesen, zeigt TextBridge zunächst das Dialogfeld "Öffnen" an (siehe Abbildung 3–6). Geben Sie dort den/die Name(n) der TIFF-Datei(en), das aktuelle Laufwerk und Verzeichnis an, und klicken Sie auf "OK".

   TextBridge öffnet das Vorschaufenster und zeigt das erste gescannte oder Online-Seitenbild an (siehe Abbildung 3–8).

5. **Wahlweise: Vergrößern Sie das Seitenbild.**

   Die Seite erscheint zunächst verkleinert, um in das Fenster zu passen. Um sie auf die volle Auflösung zu vergrößern, können Sie das Menü "Ansicht" öffnen und den Befehl **Vergrößern** wählen. Um das Seitenbild stufenweise zu vergrößern, klicken Sie auf das Vergrößerungssymbol aus der Vorschau-Toolbox, und klicken Sie auf die Seitenanzeige.

   Wenn die Seite vergrößert ist, werden Scroll-Leisten angezeigt, mit denen Sie die Anzeige horizontal und vertikal verschieben können (Abbildung 3–9).

   Zur Verkleinerung öffnen Sie wieder das Menü "Ansicht" und wählen den Befehl **Verkleinern**. Zur stufenweisen Verkleinerung klicken Sie auf das Verkleinerungssymbol und anschließend auf die Seitenanzeige.

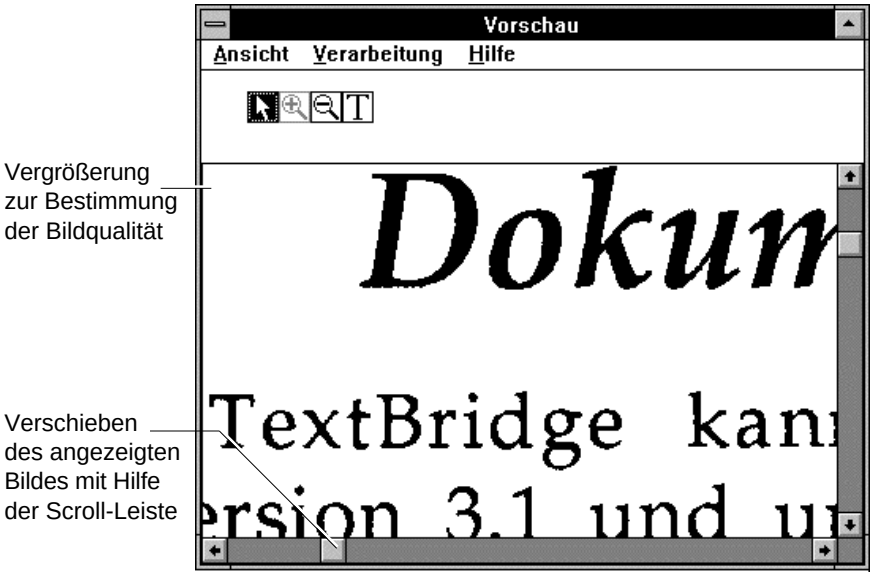

Abbildung 3–9. Vergrößerung bei der Seitenvorschau

6. **Wahlweise: Erstellen Sie einen oder mehrere zu verarbeitende Seitenbereiche.**

   Auf folgende Weise können Sie bis zu 127 Bereiche erstellen:

   - Klicken Sie auf das Textbereich-Tool, zeigen Sie damit auf eine Ecke des zu erstellenden Bereichs, klicken Sie darauf, halten Sie die linke Maustaste gedrückt und ziehen Sie die Maus. Während Sie die Maus bewegen, wird ein rechteckiger Bereich angezeigt.

   - Wenn der Bereich Ihren Wünschen entspricht, lassen Sie die Maus los.

+ Die Nummer des Bereichs erscheint in der linken oberen Ecke. Ziehpunkte erscheinen auf dem Rechteck um den Bereich herum (Abbildung 3–10).

  Die Bereichsnummer zeigt die Reihenfolge an, in der die erkannten Textblöcke in der fertigen Textdatei ausgegeben werden.

  Um die **Größe** eines Bereichs **einzustellen**, klicken Sie auf einen Ziehpunkt, halten die Maustaste gedrückt und ziehen die Maus.

  Um einen Bereich zu **verschieben**, klicken Sie auf den Bereich, halten die Maustaste gedrückt und ziehen die Maus.

  Um einen Bereich in Bezug auf einen anderen Bereich zu **positionieren**, wählen Sie den Bereich aus (Ziehpunkte müssen schwarz sein), öffnen Sie das Menü “Ansicht” und wählen den Befehl **Nach vorn bewegen** oder **Nach hinten bewegen**.

  Überlappungsbereiche sind undurchsichtig; der mehreren Bereichen gemeinsame Bildbereich ist eine Ausgabe und gehört zum jeweils oberen Bereich.

  Um einen Bereich zu **löschen**, öffnen Sie das Menü “Ansicht”, und wählen Sie den Befehl “Bereich löschen” (oder drücken Sie einfach die Löschtaste).

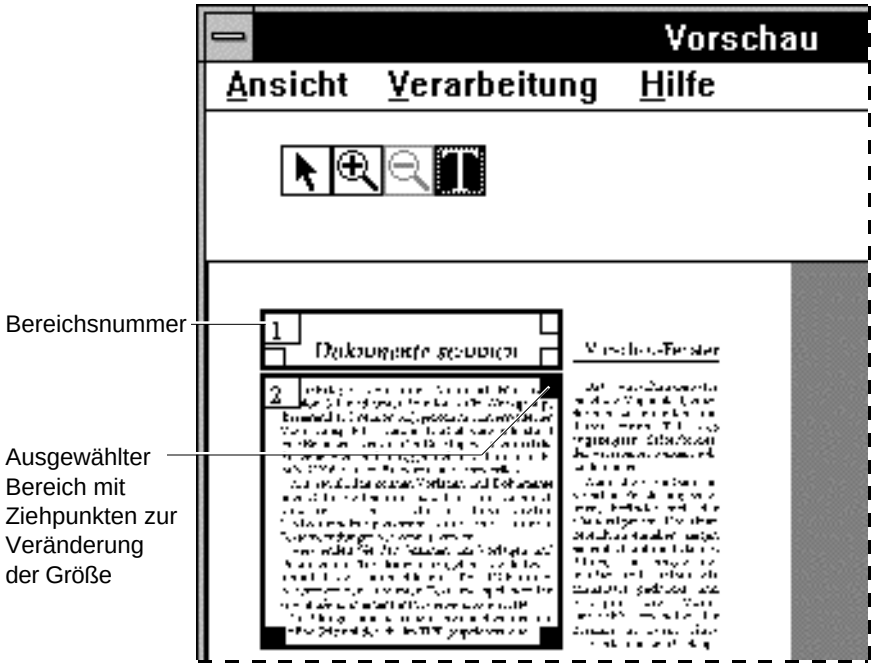

Abbildung 3–10. Bereiche bei der Vorschau

7. **Wenn Sie mit der Seitenvorschau fertig sind, beginnen Sie mit der OCR.**

   Öffnen Sie das Menü “Verarbeitung”, und wählen Sie “Diese Seite” oder “Alle Seiten”, um mit der OCR zu beginnen.

   + Wählen Sie “Alle Seiten”, um das Vorschaufenster zu schließen und alle Seiten entsprechend der festgelegten Bereiche zu verarbeiten.

     Wählen Sie “Diese Seite”, um jede Seite einzeln anzuzeigen und zu verarbeiten. TextBridge verarbeitet die erste Seite, scannt oder liest sie und zeigt die nächste Seite im Vorschaufenster an. Anschließend befindet sich TextBridge im Ruhemodus. Fahren Sie mit Schritt 5 fort, um eine neue Seite anzuzeigen.

+ Nachdem Sie die letzte Seite der Vorlage angezeigt haben, wählen Sie den Befehl “Alle Seiten”, um das Vorschaufenster zu schließen. TextBridge beendet nun die OCR der Vorlage.

Anschließend wird das Dialogfeld “Speichern unter” angezeigt (siehe Abbildung 3–4).

8. Geben Sie Informationen zur Ausgabedatei an.

- Geben Sie den Dateinamen, das Ziellaufwerk und Zielverzeichnis an.
- Bestimmen Sie auch das Ausgabeformat in dem Popup-Menü “Dateiformat”.
- Klicken Sie auf “OK”.

TextBridge konvertiert den erkannten Text in das angegebene Format, speichert ihn in der angegebenen Datei und schließt das Dialogfeld. Das Hauptdialogfeld bleibt geöffnet und ist bereit für die nächste Aufgabe.

PRÜFEN VON TEXT WÄHREND DER ERKENNUNG

Zum Erzielen der höchsten OCR-Genauigkeit – auch bei problematischen Vorlagen – verfügt TextBridge über eine Funktion zur **Prüfung** (Abbildung 3–11).

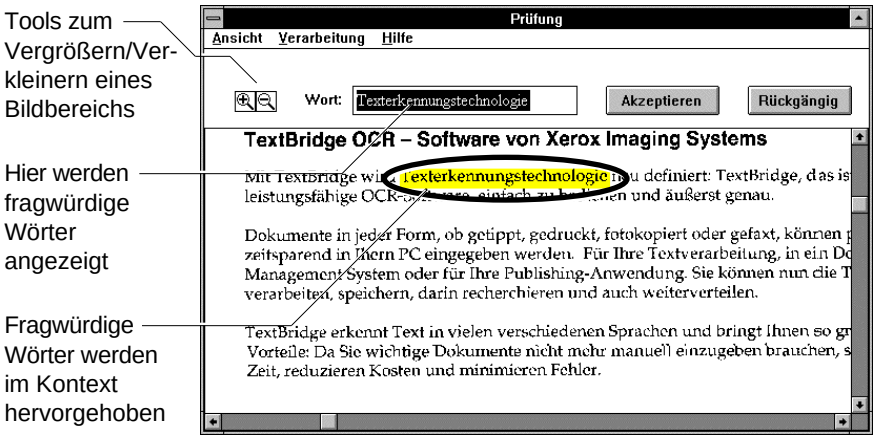

Abbildung 3–11. Prüfungsfenster

Die Prüfung zeigt Ihnen **fragwürdige Wörter** an, die Sie während der Erkennung prüfen und/oder korrigieren können.

Fragwürdige Wörter liegen unterhalb einer in TextBridge eingebauten **Zuverlässigkeitsgrenze**. Während der OCR weist TextBridge einem jeden Wort einen Zuverlässigkeitswert zu. Wenn dieser Wert unterhalb einer bestimmten Grenze liegt und Sie die Prüfung verwenden, wird das betreffende Wort als fragwürdig angezeigt.

Indem Sie Fehler korrigieren und korrekte Wörter bestätigen, helfen Sie TextBridge, die Erkennungsgenauigkeit für den Rest der Vorlage zu verbessern. TextBridge verwendet nämlich Ihre Eingaben bei den weiteren Erkennungsentscheidungen.

Das in Abbildung 3–11 gezeigte Prüfungsfenster ähnelt dem Vorschaufenster. In einem Teil des Prüfungsfensters zeigt TextBridge ein vergrößertes Bild der Seite an, die das fragwürdige Wort enthält. Das entsprechende Wort ist hervor-gehoben in seinem Kontext zu sehen.

Über dem Bilbereich finden Sie Tools zum Vergrößern/Verkleinern des Seitenbildes:

Vergrößern

Verkleinern

Neben der Toolbox befindet sich ein **Bearbeitungs-feld** mit dem erkannten Wort, welches zur Korrektur oder Bestätigung hervorgehoben ist. Rechts neben dem Bearbeitungsfeld befinden sich die Schaltflächen “Akzeptieren” und “Rückgängig”.

Im folgenden wird die Überprüfung von Wörtern während der OCR beschrieben.

Hinweis Es wird hier davon ausgegangen, daß der Scanner ordnungsgemäß an Ihren PC angeschlossen und eingeschaltet ist und daß weiterhin TextBridge installiert und gestartet ist.

---

1. **Wenn Sie scannen wollen, legen Sie die Seite(n) in den Scanner, und fahren Sie mit Schritt 2 fort. Andernfalls beginnen Sie mit Schritt 2.**

2. **Geben Sie im Hauptdialogfeld die Eingabequelle an.**

   - Wählen Sie im Dialogfeld “Eingabe aus” entweder “Datei” oder “Scanner” aus (siehe Abbildungen 3–2 und 3–5).

   - Klicken Sie im Hauptdialogfeld auf das Kontrollkästchen “Prüfung”.

3. **Wahlweise können Sie das Scannen und die OCR genauer definieren, indem Sie auf die Schaltfläche “Voreinstellungen” klicken.**

   Informationen dazu finden Sie in Tabelle 3–1. Wenn Sie fertig sind, klicken Sie auf “OK”, um zum Hauptdialogfeld zurückzukehren.

4. **Klicken Sie im Hauptdialogfeld auf “Starten!”.**

   Wenn Sie einen TWAIN-Scanner verwenden, wird die native Benutzerschnittstelle angezeigt, von der aus Sie das Scannen durchführen können. Beim ISIS-Scanner beginnt TextBridge automatisch mit dem Scannen.

   Wenn Sie die Seitenbilder von einer oder mehreren TIFF-Dateien lesen, zeigt TextBridge zunächst das Dialogfeld “Öffnen” an (siehe Abbildung 3–6). Geben Sie dort den/die Name(n) der TIFF-Datei(en), das aktuelle Laufwerk und Verzeichnis an, und klicken Sie dann auf “OK”.

   TextBridge beginnt nun mit der OCR. Wenn das erste fragwürdige Wort auftaucht, wird das Prüfungsfenster geöffnet und das fragwür-dige Wort hervorgehoben in einem Bearbei-tungsfeld angezeigt (siehe Abbildung 3–11). Unterhalb des Bearbeitungsfeldes zeigt das Prüfungsfenster das Wort hervorgehoben in seinem Kontext an

5. **Prüfen und/oder korrigieren Sie das fragwürdige Wort im Bearbeitungsfeld, und klicken Sie auf die Schaltfläche “Akzeptieren”.**

   + Wenn Sie während des Prüfvorgangs einen Fehler machen, klicken Sie einfach auf die Schaltfläche **Rückgängig**. Das Bearbeitungs-feld stellt das letzte von Ihnen editierte Wort wieder her, und Sie können den Fehler korrigieren.

+ Sie können die Empfindlichkeit der Prüffunktion mit dem Befehl **Prüfung** und seinen fünf möglichen Einstellungen steuern. Öffnen Sie das Menü “Verarbeitung”, wählen Sie dann den Befehl “Prüfung” und anschließend die Einstellung ”Maximal”, “Mehr”, “Normal”, “Weniger” oder “Minimal”. Die Einstellungen “Mehr” und “Maximal” haben zur Folge, daß mehr Wörter zur Überprüfung angezeigt werden. “Weniger” oder “Minimal” zeigen weniger Wörter an. “Normal” ist die Standard-einstellung.

Manche Vorlagen enthalten Wörter, deren Buchstaben nicht auf Ihrer Tastatur vorhanden sind (Akzente, Symbole, etc.). Zum Überprüfen solcher Zeichen öffnen Sie das Menü “Ansicht” und wählen Sie **Sonderzeichen anzeigen**. Eine Tastatur mit Sonderzeichen wird angezeigt. Um diese ins Bearbeitungsfeld einzugeben, wählen Sie sie auf der angezeigten Tastatur aus, indem Sie mit der Maus darauf klicken (Abbildung 3–12).

Eingabe eines Sonderzeichens durch Anklicken

Abbildung 3–12. Sonderzeichen-Tastatur

+ Hin und wieder hält TextBridge **Störfaktoren** (bestimmte Zeichen auf der Seite) oder eine horizontale Linie für Text. Im Prüfungsfenster enthält das Bearbei-tungsfeld Zeichen, obwohl im Bildbereich ein “Nicht-Wort” hervorgehoben ist. Löschen Sie in solchen Fällen den gesamten Text im Bearbeitungsfeld, und klicken Sie auf “Akzeptieren”. TextBridge ignoriert dann die Störfaktoren und fährt mit den fragwürdigen Wörtern fort.

  Wenn das Erscheinungsbild eines bestimmten Wortes im Vergleich zum Rest der Vorlage schlecht ist, sollten Sie den Text korrigieren, ohne TextBridge das Bild “beizubringen”. Korrigieren Sie das Wort einfach im Bearbeitungsfeld, halten Sie die Steuerungstaste gedrückt und drücken Sie auf die Eingabetaste (oder klicken Sie auf “Akzeptieren”). Der korrigierte Text wird entsprechend ausgegeben.

  Der Bilbereich im Prüfungsfenster wird gemäß einer mittleren Einstellung vergrößert. Wenn Sie Genaueres über die Wortposition auf der Seite erfahren wollen, klicken Sie auf das Verkleinerungssymbol und klicken Sie den zu prüfenden Bildbereich an. Das gesamte Seitenbild wird im Bildbereich des Prüfungsfensters angezeigt. Wenn Sie die Anzeige vergrößern wollen, klicken Sie auf das Vergrößerungssymbol.

6. Wiederholen Sie Schritt 5 solange, bis genügend Wörter geprüft wurden.

+ Prüfen Sie mindestens eine Seite einer mehrseitigen Vorlage, um dem System genügend “Wissen” über die gesamte Vorlage beizubringen.

7. **Schließen Sie die Prüffunktion.**

   Öffnen Sie das Menü “Verarbeiten” und wählen Sie den Befehl **Prüfung beenden**.

   + Der Befehl “Schließen” im Steuerungsmenü des Prüfungsfensters entspricht dem Befehl “Prüfung beenden”.

   Das Prüfungsfenster wird geschlossen, und die OCR wird automatisch fortgeführt. Wenn Sie mit der Verarbeitung aller Seiten fertig sind, zeigt TextBridge das Dialogfeld “Speichern unter” an (siehe Abbildung 3–4).

8. **Geben Sie die Ausgabeinformationen an.**

   - Bestimmen Sie den Dateinamen, das Ziellaufwerk und -verzeichnis.
   - Geben Sie auch das Ausgabeformat in dem Popup-Menü “Dateiformat” an.
   - Klicken Sie auf “OK”.

   TextBridge konvertiert den erkannten Text in das angegebene Format, speichert ihn in der angegebenen Datei und schließt das Dialogfeld. Das Hauptdialogfeld bleibt geöffnet und TextBridge ist bereit für die nächste Aufgabe.