

# Paper Capture in der Praxis

## Wie kann ich gescannte Texte in durchsuchbare PDF-Dateien umwandeln?

Acrobat bietet die Möglichkeit, Papierdokumente direkt beim Scannen in PDF-Dokumente umzuwandeln und durch die Acrobat-eigene OCR-Funktion gleichzeitig durchsuchbaren Text zu generieren. Was aber ist mit PDF-Dokumenten, die lediglich Bitmap-Daten beinhalten, womit die darin enthaltenen Texte nicht durchsuchbar sind? Abgesehen davon, dass gescannte Text-PDF-Dokumente mehr Speicherplatz als reine Text-PDF-Dokumente benötigen, können die gescannten Daten nicht editiert, in andere Programme kopiert, durchsucht oder indiziert werden.

Acrobat bietet über das Plug-in PAPER CAPTURE die Möglichkeit, gescannte Texte in editierbare Texte zu konvertieren. Diese Texte können dann korrigiert, indiziert, durchsucht und in andere Dateien kopiert werden. Bei dem Plug-in PAPER CAPTURE handelt es sich um eine OCR-Software (Optical Character Recognition), die gescannten Text erkennt und in editierbaren bzw. durchsuchbaren Text umwandelt. Das Ergebnis dieser Umwandlung hängt jedoch stark von der Qualität der gescannten Vorlage ab. Im Idealfall werden alle Zeichen korrekt erkannt und umgewandelt. ACRBAT PRO/PRO EXTENDED bietet die Möglichkeit, fehlerhaft erkannte Texte nachträglich einfach zu korrigieren.

Wenn Sie ein gescanntes Dokument über PAPER CAPTURE in editierbaren Text umwandeln, kann die Umwandlung so durchgeführt werden, dass anschließend nur noch der umgewandelte Text vorhanden ist, es besteht aber auch die Option, den umgewandelten Text auf eine eigene Ebene unterhalb des gescannten Dokuments zu platzieren. Dadurch bleibt das optische Erscheinungsbild erhalten, der Text jedoch kann durchsucht werden.

## Wie kann ich Text in einem gescannten PDF-Dokument in eine durchsuchbare PDF-Datei umwandeln?

1. Öffnen Sie das PDF-Dokument, das gescannten Text enthält, den Sie über die Acrobat-eigene OCR-Software in durchsuchbaren Text umwandeln möch-



### 9 Versionsinfo

Die Beschreibungen wurden anhand der Version ACRBAT PRO 9 erstellt.

### PDF beim Scannen erzeugen

#### Verweis

Nähere Informationen zum Thema vom Scanner direkt zum PDF erhalten Sie im Heft PDF PUBLIZIEREN im Kapitel PDF-ERZEUGUNG IN ACRBAT.

### Paper Capture

#### Hinweis

Um die OCR-Texterkennung auf ein PDF-Dokument anwenden zu können, muss die ursprüngliche Scannerauflösung mindestens 72 dpi betragen haben.  
Das Scannen mit 300 ppi (Pixel pro Zoll) führt zu den besten Ergebnissen. Bei 150 ppi (Pixel pro Zoll) ist die OCR-Erkennung etwas ungenauer.

### Zwei Arten der Texterkennung

#### Texterkennung starten

#### CD-ROM

Auf der beigefügten CD finden Sie im Verzeichnis ARBEITSDATEIEN\PROFI-KNOW-HOW\PAPERCAPTURE Beispieldateien, die Sie verwenden können, um die beschriebenen Arbeitsschritte nachzuvollziehen.

Der Befehl zum Starten der Texterkennung wird über den Menübefehl »Dokument → OCR-Texterkennung → Text mit OCR erkennen« ausgeführt.