

Textauszug aus Papierdokument übernehmen

Guter Ausdruck vorteilhaft

! Hinweis

OCR (Optical Character Recognition) oder »Optische Zeichenerkennung« ist ein Verfahren zur automatischen Texterkennung einer gedruckten Vorlage.

Texte extrahieren

Schnelle Lösung ►

Ausführliche Lösung ►

Texte und Bilder aus Papierdokumenten übernehmen

Kann ich Text aus einem Dokument extrahieren, das nur auf Papier und nicht digital vorliegt?

Es ist gar nicht so selten, dass Ihr Kunde Ihnen bestimmte Texte und Dokumente nur auf Papier liefern kann, sei es, dass die Originaldateien verloren gingen oder dass das Material uralt ist und in einem längst vergessenen Programm erzeugt wurde oder sogar noch nie in digitaler Form vorlag.

Wenn Sie Adobe Acrobat besitzen, stellt dies zumindest bezüglich der Texte kein allzu großes Problem für Sie dar, vorausgesetzt, dass diese in einer klaren Schrift abgedruckt sind. Bei geringer Druckqualität, verschmutzten Seiten und vom Standard abweichenden Schriften funktioniert das Verfahren zwar auch, jedoch ist der Nachbearbeitungsbedarf höher, weil wahrscheinlich viele Buchstaben nicht richtig erkannt werden.

Das Programm enthält eine Funktion namens »Paper Capture«, mit der Sie Papierdokumente in Acrobat einscannen und per OCR in editierbaren Text konvertieren können. Das Besondere dabei ist, dass die Schriftzeichen in diesen gescannten Dokumenten in »richtigen« Text umgewandelt werden, den Sie anschließend bearbeiten können.

Für eine erfolgreiche Durchführung der im Folgenden beschriebenen Schritte benötigen Sie einen Scanner (oder die eingescannten Dokumente) und eine ganze Palette der Creative-Suite-Programme: Adobe Acrobat, Photoshop und InDesign.

Wie extrahiere ich Texte aus einem Papierdokument?



Wenn Sie das Dokument als (Aus-)Druck auf Papier oder Film erhalten, benötigen Sie zunächst einen Scanner.