

«DURCHKLICK»: VOM PDF ZUM DURCHSUCHBAREN DOKUMENT BZW. WORD-DOKUMENT

ADRIAN RUFENER

lic. iur., Rechtsanwalt, St. Gallen

Stichworte: PDF, Scanning, OCR, Texterkennung, Volltextsuche, Word

PDF hat sich als Standard für den plattformunabhängigen Abruf und Transport von Dokumenten durchgesetzt. Mit der «richtigen» Nachbearbeitung sind gescannte PDF-Dokumente im Volltext durchsuchbar bzw. können in der Office-Umgebung als Word oder Excel weiterverarbeitet werden.

I. Ausgangslage

1. PDF als Standard für die Übermittlung von Dateien

Das Portable Document Format ist ein plattformunabhängiges Dateiformat für Dokumente, das vom Unternehmen «Adobe Systems»¹ entwickelt und 1993 veröffentlicht wurde. Ziel war es, ein Dateiformat für elektronische Dokumente zu schaffen, das diese unabhängig vom ursprünglichen Anwendungsprogramm, vom Betriebssystem oder von der Hardwareplattform originalgetreu weitergeben kann. Ein Leser einer PDF-Datei soll das Dokument immer in der Form betrachten und ausdrucken können, die der Autor festgelegt hat. Die typischen Konvertierungsprobleme (wie veränderter Seitenumbruch oder falsche Schriftarten) beim Austausch eines Dokuments zwischen verschiedenen Programmen entfallen dadurch.

PDF hat sich inzwischen zum Standard entwickelt, Dokumente einer beliebigen Anzahl von Benutzern zur Verfügung zu stellen, sei es als Gesetzestexte, als Bedienungsanleitungen, als Dokumente im Austausch mit Klienten, im elektronischen Rechtsverkehr, als ausfüllbare Formulare etc.

2. PDF erstellen

Auf dem Markt sind verschiedene Softwarelösungen², auch kostenlose³, erhältlich, mit welchen Dokumente bzw. Inhalte in ein PDF-Dokument «gedruckt» bzw. umgewandelt werden können. Dokumente, welche mit den «PDF-Druckern» erstellt worden sind, lassen sich mit «PDF-Readern»⁴ lesen und im Volltext durchsuchen.

3. Multifunktionsgeräte/Scanner

Marktgängige Scanner und/oder Multifunktionsgeräte⁵ ermöglichen es, Dokumente nicht bloss im sonst üblichen TIFF-Format⁶ zu scannen, sondern auch als PDF zu erstellen. Das Besondere an gescannten Dokumenten besteht

darin, dass die Scanner ein Abbild des elektronisch «fotografierten Bildes» erstellen. Gescannte Dokumente sind als «Bilder» ohne Nachbearbeitung nicht durchsuchbar und eignen sich deshalb nicht zur Ablage in Datenbanken bzw. in der Mandatsverwaltungssoftware.

II. Nachbearbeitung eines gescannten PDF-Dokumentes

1. Nicht durchsuchbares «Bild» aus Ausgangspunkt

Gescannte Dokumente, sei es ein elektronisches Dossier einer IV-Stelle, welches auf einer CD als PDF zur Verfügung gestellt wird, oder ein «alter Vertrag», welcher von einem Klienten zur Überarbeitung zugestellt wird, lassen sich einerseits einfach volltextindexieren und andererseits direkt in Word umwandeln.

2. Volltexterkennung

Mittels OCR⁷-fähiger Software⁸ lassen sich Texte innerhalb von «Bildern» automatisch erkennen. In Sekundenschnelle verwandelt die Software nicht durchsuchbare Bilder in PDF-Dokumente, welche im Volltext durchsucht werden können. Ein weiterer Vorteil durchsuchbarer Dokumente besteht darin, dass Texte mittels «copy and

1 <http://www.adobe.com/de/>.

2 Acrobat XI Pro.

3 Z. B. «PDF 24 creator» unter: <http://de.pdf24.org/>.

4 Z. B. Adobe Reader XI (kostenlos erhältlich unter: <http://get.adobe.com/de/reader/>).

5 Mit den Funktionen: Drucken, Scannen, Kopieren, Faxen etc.

6 Tagged Image File Format.

7 Optical Character Recognition.

8 Z. B. Adobe Acrobat XI Pro; OmniPage 18 (unter: www.nuance.de).

paste» in die gängigen Office-Programme übernommen werden können.

3. *Umwandlung in Word*

Mit Adobe XI Pro, OmniPage oder weiteren OCR-fähigen Programmen können sodann PDF-Dokumente direkt in Word-Dokumente umgewandelt und anschliessend nachbearbeitet werden. Es muss wohl nicht näher dargelegt werden, dass der Nutzen einer solchen Möglichkeit gross ist und die Beschaffungskosten⁹ der entsprechenden Software bei Weitem rechtfertigt.

III. Und was Adobe XI Pro u. a. sonst noch kann

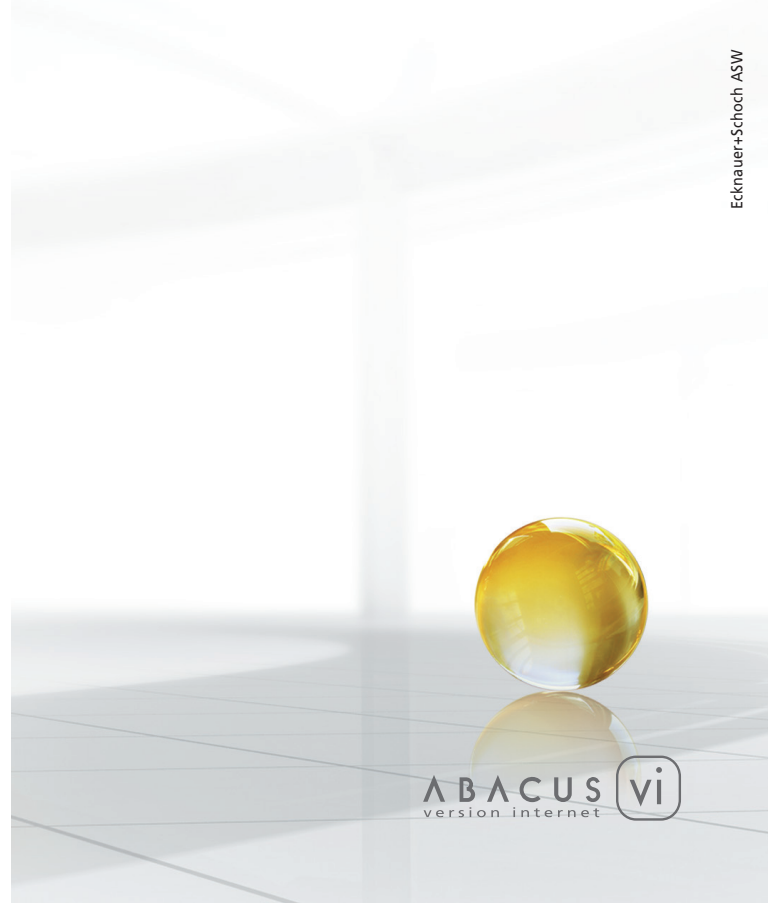
1. *Seiten einfügen und entnehmen*

Oftmals ist es hilfreich, einer Gegenpartei oder einem Klienten Textausschnitte, Teile eines Dokumentes oder Ausschnitte aus Gesetzestexten zur Verfügung zu stellen. Mit Adobe XI Pro können unter anderem Seiten aus PDF-Dokumenten entnommen oder gelöscht werden. Sodann können andere PDF-Dokumente seitenweise oder vollständig eingefügt werden.

2. *Texte gelb markieren/schwärzen*

Im Kanzleialltag ist es auch hilfreich, wenn Textausschnitte in PDF-Dokumenten mit der Markerfunktion gelb markiert werden können. Sodann können mit Adobe XI Pro Textpassagen und/oder Begriffe geschwärzt werden. Dabei lässt es die Software zu, im Text nach Begriffen zu suchen und diese zu schwärzen.

⁹ OmniPage 18 kostet ca. CHF 250.- pro Arbeitsplatz; Adobe XI Pro kostet knapp CHF 800.- und kann auf zwei Rechnern installiert werden. Die Adobe-XI-Standardversion ist für weniger als CHF 500.- erhältlich.



Business Software für rationelle Leistungserfassung

- > Flexible Definition von Leistungsarten
- > Freies Customizing der Mandatsstammdaten
- > Erfassung von Stunden, Drittleistungen, Spesen, Absenzen
- > Web-Erfassung
- > Stundenkontrolle nach verrechenbaren Stunden, Gleitzeit, Absenzen
- > Fristen- und Aktivitätenkontrolle
- > Projektübersichten mit Auftrags-eingängen und Projektabschlüssen
- > Automatische Fakturierung von Pauschalen, Vertragshonoraren
- > Produktivitätsauswertungen

www.abacus.ch

 **ABACUS**
business software