



PDF-Verarbeitung in der Praxis Adobe® Acrobat® XI

INHALT

1	Dokumente im PDF-Format	1
1.1	Preis- und Aufwandskalkulationen auf PDF-Basis	2
1.2	Direktes Bearbeiten/Editieren des PDF	2
1.3	Generierung einer editierbaren Datei aus einem PDF	3
2	Ergebnisse mit Adobe® Acrobat® XI	4
2.1	PDF direkt als CAT-Projekt (hier MemoQ) einlesen	4
2.2	PDF Version XI (erstellt z. B. aus Word oder InDesign) speichern nach MS-Word...4	
2.3	PDF mit nicht anfassbarem Text (Grafik), OCR-Erkennung und Speichern nach MS-Word.....	5
3	Fazit	6
3.1	Handhabung und Aufwand.....	6
4	Allgemeine Informationen und Tipps zu PDF.....	7
4.1	Kommentieren von PDF	7

1 Dokumente im PDF-Format

Das Dateiformat PDF ist heutzutage ein nicht mehr wegzudenkender Standard für die Verteilung von Dokumenten.

Was zunächst – und in diesem Rahmen des Dateiaustausches – als ein flexibles, system- und programmunabhängiges Format gilt, hat für die Weiterverarbeitung, wie sie in der Übersetzungsbranche erforderlich ist, immer noch so seine Tücken.



1.1 Preis- und Aufwandskalkulationen auf PDF-Basis

Für eine einfache Textkalkulation, d.h. die Ermittlung der Wortanzahl, wird das PDF direkt eingelesen und analysiert. Dieses setzt voraus, dass das PDF ursprünglich ohne Sondereinstellungen, ohne Passwortschutz, ohne Druckschutz und ohne Legenden als Grafiken erstellt worden ist.

1.2 Direktes Bearbeiten/Editieren des PDF

Immer noch gibt es die weit verbreitete Meinung, man könne direkt im PDF die Texte überschreiben. Stimmt: über „Erweiterte Bearbeitung“ und „Touch-Up Textwerkzeug“ kann man Worte einer Zeile im PDF editieren und mit dem neuen Text ersetzen. Kommt man dann ans Ende der Zeile, schließt man diese ab und setzt seinen Text in der nächsten Zeile fort. Legt man keinen großen Wert auf flüssige Textinhalte und hat zudem viel Zeit, geht das schon irgendwie. Folgendes gilt es aber zu berücksichtigen:

- Die im PDF verwendete Systemschrift muss vorliegen oder muss gegen eine entsprechende ersetzt werden können
- Der neue Textinhalt darf nicht expandieren, da nur ein Gesamtextrahmen zur Verfügung steht und das zu einen Überlauf auf die Folgeseite führen kann (siehe unten)
- Die Zeilenumbrüche sind manuell zu gestalten
- Rechtschreibprüfung gibt es nicht
- Bestimmte Steuerzeichen wie Tabulator funktionieren nicht
- Tabellen sind nicht editierbar bzw. „zerlegen“ sich
- Formate werden zum Teil zerstört (siehe unten)

Summe (Rest) **Formate werden zerstört**

In diesem Beispiel lassen Sie eine Übersetzung ins Englische (Punkte = 20.300 x 1 x 2/100), eine Übersetzung ins Polnische (Punkte = 800 x 0,9 x 2/100) und eine Übersetzung ins Französische machen, sammeln so also 580 Punkte. Bei der Übersetzung am 15. Mai, für die Sie wiederum 24 Punkte erhalten, setzen Sie einen Teil der Gesamtpunkte kostenreduzierend ein und brauchen statt 1.200 Worte nur 650 Worte (abzüglich 550 x 1) zahlen.

Weitere Fragen zu diesem Punktesystem beantworten wir Ihnen gerne unter +49 (0) 40 409409 oder über eine E-Mail an pool@tts-europe.de.

Überlauf auf die Folgeseite

TTS GmbH



1.3 Generierung einer editierbaren Datei aus einem PDF

Zwar war und ist es möglich mit Acrobat® 8 den Text aus dem PDF-Format in ein anderes – wie z.B. Word – zu bringen, doch wie so häufig liegt der Teufel im Detail. Textinhalte fanden sich in Textfeldern oder gar Positionsrahmen wieder, jede Seite begann mit neuem Abschnittswechsel und Absatz- wie auch Schriftformate waren eigenständig geändert. Insbesondere unter Verwendung von Translation Tools wie beispielsweise memoQ oder Trados ist bei der weiteren Bearbeitung für die Übersetzung aber Vorsicht geboten, da die Texte nicht in einer logischen Reihenfolge übernommen werden. Das bedeutet, je komplizierter der Text aufgebaut ist – z.B. über mehrere Spalten, viele Tabellen etc. – desto mehr Vorarbeit ist notwendig, um eine richtige und überschaubare Textreihenfolge zu erlangen. Im ungünstigsten Fall muss jeder Satz kontrolliert und evtl. an einen anderen Platz gestellt werden.

Für reinen Fließtext mit ein paar Einrückungen reicht die Funktion „Datei in Zwischenablage kopieren“ vollständig aus (hierbei müssen allerdings die Absatzmarken am Ende jeder Zeile entfernt werden).

So mussten wir immer wieder auf OCR-Anwendungen wie ABBYY FineReader zurückgreifen, um ein weiter zu verarbeitendes Format zum Übersetzen zu erhalten. Doch auch bei OCR-Erkennung muss das Layout zumeist komplett überarbeitet werden.

Oftmals bietet es sich dann an, den Text im Tabellenformat in MS-Word zu erfassen, wenn Sie als Kunde nur die Übersetzung benötigen und das Layout in einer anderen Anwendung wie z.B. QuarkXpress oder InDesign erstellen. Durch diese Methode kann einigermaßen sichergestellt werden, dass der Layouter die richtige Übersetzung für die jeweilige Textpassage übernimmt. Ein Endlektorat des Druck-PDF ist dennoch bei dieser Methode unabdingbar.

Mit Adobe® Acrobat® X in 2011 sollte alles besser werden, die Hoffnung war aber nach ersten Versuchen schnell dahin.

Die aktuelle Version Adobe® Acrobat® XI verspricht nunmehr: *„Mit Acrobat XI können Sie Ihre PDF-Dokumente mit allen Formatierungen als Microsoft Word-Dokument speichern und leicht bearbeiten“*. Gleiches soll für Excel und PowerPoint gelten.

Das wollten wir genau wissen und haben unseren Expertinnen eine Testversion installiert.



2 Ergebnisse mit Adobe® Acrobat® XI

Vorbemerkung: Das Ergebnis hängt sehr davon ab, wie stark das PDF formatiert ist, wie komplex sein Layout ist.

Für die Tests wurde versucht, Dateien zu verwenden, die stärker formatiert und umfangreicher sind.

2.1 PDF direkt als CAT-Projekt (hier MemoQ) einlesen

- Bei nicht stark formatierten Texten: bedingt brauchbar
- Bei stark formatierten Texten: ohne genaue Prüfung auf korrekte Segmentierung und Vollständigkeit nicht brauchbar

2.2 PDF Version XI (erstellt z. B. aus Word oder InDesign) speichern nach MS-Word

Was positiv auffällt:

- Im Gegensatz zu älteren Acrobat®-Versionen fällt positiv auf, dass nicht mehr alles, was eingerückt ist oder z.B. vom linksbündigen Textfluss abweicht (wie Legenden bei Abbildungen etc.) in separate Textfelder verteilt wird (diese lagen häufig auch noch übereinander).
- Anstelle der Textfelder wird mit Absatzeinrückungen gearbeitet, sehr viel mit Abschnittswechseln und Spaltenumbrüchen. Auch ein großes 200-Seiten-PDF wurde nach Word konvertiert und sah dem Original durchaus ähnlich.
- Sehr hilfreich ist, dass alle Objekte bei der Inhaltsbearbeitung separat anzufassen sind, d.h. alle Textblöcke und Bilder können vor dem Export, falls erforderlich, bearbeitet, ergänzt oder auch gelöscht werden. Das macht es theoretisch möglich, den übersetzten Text direkt ins PDF zu kopieren, allerdings Textblock für Textblock mit individueller Anpassung, evtl. treten Probleme wg. fehlender Schriften auf. Dieses Vorgehen ist aber sehr zeitintensiv!



Was negativ auffällt:

- Kopfzeilen wurden nicht vollständig erkannt
- Diverse nicht erforderliche Abschnittswchsel statt einfacher Seitenumbrüche
- Nicht erforderliche Absatzmarken am Zeilenende (mitten im Satz)
- Einfache Tabellen werden korrekt übernommen, sobald die Tabellen komplizierter werden, sind sie nicht sauber formatiert
- Automatische Nummerierungen funktionieren nur zum Teil
- Schriften/Schriftgrößen müssen überprüft/vereinheitlicht werden
- Bei fremdsprachigen Texten gab es einige Fehler bei Sonderzeichen, zum Teil wurden Versalien und Kapitälchen nicht korrekt übernommen
- Generell sehr viele Abschnittswchsel und Spaltenumbrüche

2.3 PDF mit nicht anfassbarem Text (Grafik), OCR-Erkennung und Speichern nach MS-Word

Funktioniert als reiner Textexport, wobei der Text allerdings sehr fehlerhaft ist (kein ß, keine Umlaute):

Nach Bearbeitung mit der OCR-Funktion ist das Ergebnis schon wesentlich besser: Bei der OCR-Bearbeitung müssen alle nicht erkannten Wörter bestätigt werden, ein bisschen mühsam, aber lohnt sich, als Grundlage zu gebrauchen.

Die Anschaffung einer getrennten Anwendung wie ABBYY FineReader (OCR) erspart man sich.



3 Fazit

Der Text kann mit der neuen Acrobat-Version leichter extrahiert werden, auch der Zugriff auf die Bilder ist einfacher.

Für die Übersetzung ist die Speicherung nach Word weiterhin erforderlich, damit der Text mit CAT-Tools bearbeitet werden kann. Vorher muss eine genaue Überprüfung der so erzeugten Word-Datei auf Vollständigkeit und korrekte Segmentierung erfolgen. Will man nach der Übersetzung eine Version beispielsweise in Word oder InDesign erzeugen, die dem Original-PDF entspricht, ist in der Regel eine umfangreiche Nacharbeit erforderlich, Anpassung wegen längerer Texte, Vereinheitlichung von Formaten, Nacharbeiten von automatischen Nummerierungen, Erstellung von Verzeichnissen etc.

3.1 Handhabung und Aufwand

Viele Anwender unterschätzen oft den Arbeitsaufwand, der mit PDF-Dokumenten verbunden ist. „Copy and Paste“ reicht leider oftmals nicht aus, um eine nicht nur inhaltlich, sondern auch optisch und sachlich qualitativ hochwertige Arbeit abliefern zu können, die Sie als Kunde zu Recht von uns verlangen.

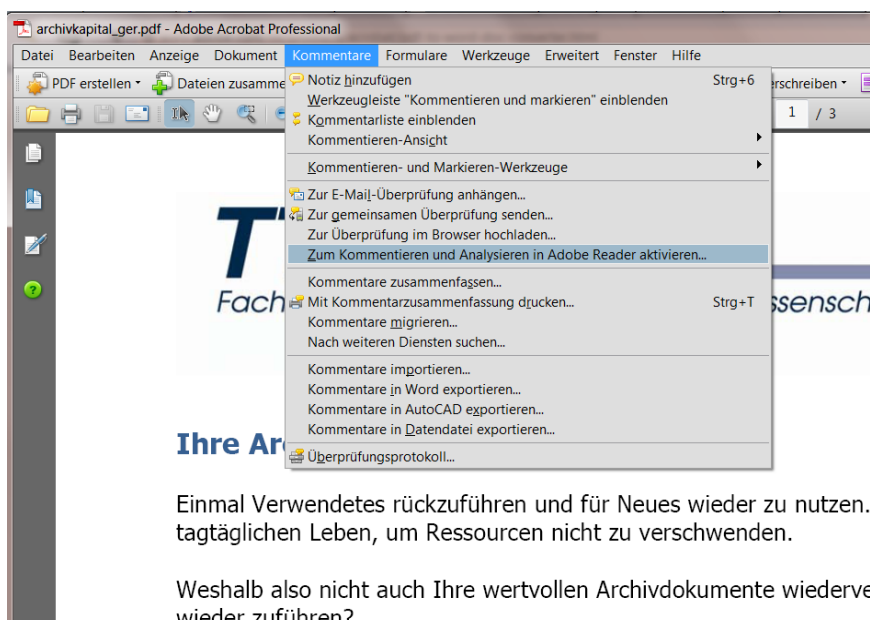
Da oft nur PDF-Formate verfügbar sind, weil Originale einfach nicht vorliegen, ist TTS selbstverständlich bemüht, diesen Umständen gerecht zu werden. Bitte haben Sie aber Verständnis dafür, dass wir nicht alle Layout-Nachbearbeitungen kostenlos durchführen können. Wir hoffen, Ihnen mit dieser kurzen Beschreibung einen kleinen Einblick in die Weiterverarbeitung von PDF-Formaten gegeben zu haben – sollten Sie weitere Informationen wünschen, rufen Sie uns gerne an.



4 Allgemeine Informationen und Tipps zu PDF

4.1 Kommentieren von PDF

Dem Leser kann die Berechtigung erteilt werden, das PDF zu kommentieren. Das ist insbesondere dann wichtig, wenn jemand lektorieren soll, und Korrekturen nicht mit missverständlichen Positionsangaben in einer Textdatei vollziehen muss. Hierzu ist die Vorgabe im Menü Kommentare auszuwählen (siehe unten). Genutzt werden sollte hierfür eine Kopie des Dokuments, da diese Einstellung zu Einschränkungen in der Nutzung des PDF führt, wie beispielsweise, dass keine Seiten mehr gelöscht werden können.



Einmal Verwendetes rückzuführen und für Neues wieder zu nutzen.
tagtäglichen Leben, um Ressourcen nicht zu verschwenden.

Weshalb also nicht auch Ihre wertvollen Archivdokumente wiederver-
wieder zuführen?