

## Merkblatt «PDF zu Word/Excel online»

### 1 Vorbemerkung

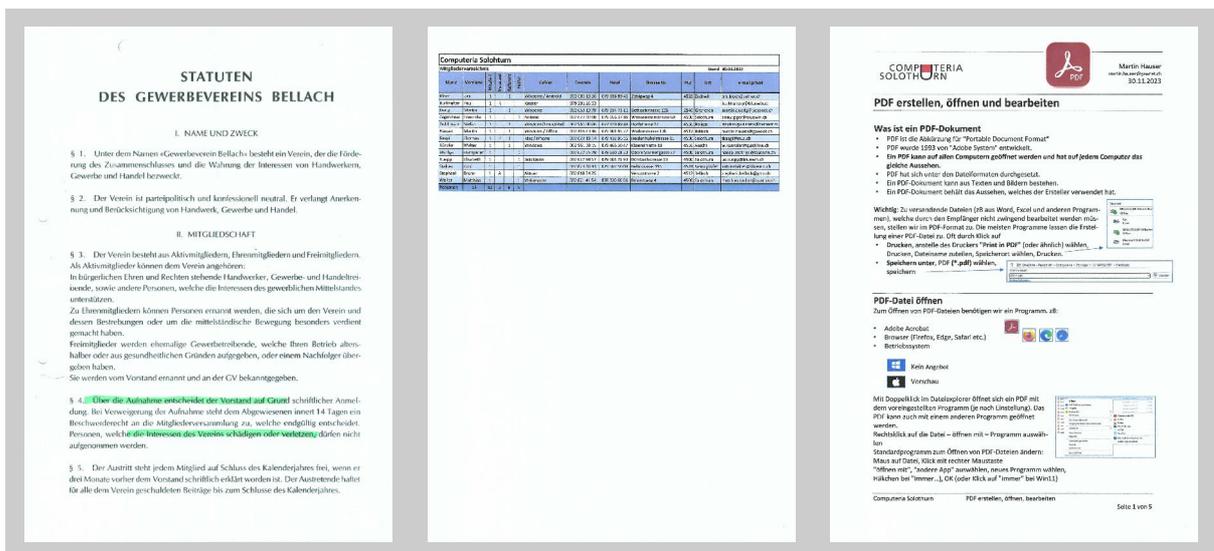
Es geht hier nur um die Konvertierung eingescannter Dokumente, die beim Scannen zwar als Bild erfasst, aber als PDF-Datei gespeichert wurden. also nicht um PDF-Dokumente, in denen der Text als solcher ausgewählt werden kann. Zur Konvertierung braucht es damit eine OCR-Funktion.

Für die folgenden Beurteilungen wurden mit einer Suche nach «Scan to Word», bzw. «Scan to Excel» gefundene Konverter miteinander verglichen. Sicher gibt es noch weitere Konverter, die qualitativ oder punkto Benutzerfreundlichkeit die aufgeführten möglicherweise sogar übertreffen. Die Beurteilung erhebt deshalb weder den Anspruch auf Vollständigkeit, noch dass die wirklich Besten gefunden wurden. Konverter, die im Vergleich deutlich abfielen, sind – abgesehen von Adobe – nicht aufgeführt.

Wurde für das Scannen des Ausgangsdokuments das Programm «Vorschau» (nur Mac) benutzt, oder ein anderes Programm, welches OCR beim Scannen unterstützt, so ist der Text im resultierenden PDF-Dokument bereits als solcher enthalten. Dann kommen zur Erzeugung eines Word- oder Excel-Dokumentes auch alle PDF-Konverter in Frage, welche die OCR-Funktion nicht unterstützen.

### 2 PDF → Word

Im Folgenden werden anhand von drei eingescannten Dokumenten ein paar Konverter verglichen, die Word-Dokumente mit Text liefern, der weiter bearbeitet werden kann. Getestet wurde mit einem reinen Text-Dokument, einer Tabelle und einem Dokument, in welchem neben dem Text verschiedene Bilder/Grafiken angeordnet sind



#### 2.1 Adobe <https://www.adobe.com/de/acrobat/online/pdf-to-word.html>

Die OCR-Funktion ist beim Text-Dokument völlig ungenügend, das Resultat unbrauchbar – sehr erstaunlich, dass ausgerechnet der Erfinder des PDF-Formates dies nicht besser beherrscht. Um mehrere Dokumente zu konvertieren, oder um konvertierte Dokumente herunterladen zu können, braucht es ein Adobe-Account, für dessen Einrichtung eine Kreditkarte angegeben werden muss. Angesichts der Ergebnisse des ersten Tests erübrigt sich dies.



Zwei gefundene Konverter fallen ausser Betracht: <https://www.online-convert.com/de> liefert bis zur Unkenntlichkeit verzerrte Tabellen und <https://www.pdfexcel.com/de> ordnet den Inhalt den Zellen nicht korrekt zu. Geprüft wurden schliesslich nur die drei bei der PDF-zu-Word-Konvertierung erfolgreichen Webseiten. Auch bei Freeconvert kann XLSX als Zielformat angegeben werden. Das Resultat ist aber eine XLSX-Arbeitsmappe, die der Tabelle des Ausgangsdokumentes überhaupt nicht entspricht. Damit bleiben nur noch zwei Kandidaten:

### 3.1 OnlineOCR <https://www.onlineocr.net>

Die reine Zahlentabelle wird perfekt umgesetzt. Bei der Adresstabelle gibt es minimale OCR-Probleme. Vor allem aber werden die Zahlengruppen der Telefonnummern in einzelne Spalten umgebrochen. Die Spalte mit der PLZ ist so schmal, dass die Ziffern nicht Platz finden und deshalb mit ## dargestellt werden.

### 3.2 Convertio <https://convertio.co/de/ocr/>

Die reine Zahlentabelle wird auch von Convertio perfekt umgesetzt. Allerdings stören die sehr kräftigen Linien des Zellgitters. Bei der Adresstabelle gibt es minimale OCR-Probleme. Die vertikal laufenden Texte des Originals werden nicht gedreht und einige Texte sind in den Zellen falsch ausgerichtet. Die Spaltenbreiten entsprechen dem eingescannten Dokument. Die Linien des Zellgitters sind auch hier viel zu kräftig.

### 3.3 Die Mac-Alternative «Vorschau»

Das für die Konvertierung zu Word durchaus geeignete «hauseigene» Mac-Programm «Vorschau» kommt hier weniger in Frage, da es keine Tabelle erzeugen kann. Handelt es sich beim verfügbaren PDF-Dokument um eines, das aus einem Tabellenbearbeitungsprogramm wie Excel exportiert wurde – also keine Bild-Datei, sondern ein «echtes» PDF –, so lässt sich dieses mit «Vorschau» öffnen. Wird danach der ganze Text in der Tabelle ausgewählt und kopiert, so lässt er sich mit «Einfügen» in eine Tabelle, z.B. von Excel, übertragen. Allerdings müssen danach die Zellen per Formel richtig angeordnet werden. Es gilt somit abzuwägen, ob der perfekt erkannte Text den Umweg wert ist.

### 3.4 Bilanz

Beide obigen online Konverter liefern brauchbare Resultate, die aber nachbearbeitet werden müssen. Die von Convertio gelieferte XLSX-Tabelle benötigt aber deutlich weniger Nachbearbeitung. Sollte mit einem anderen Dokument Convertio kein befriedigendes Resultat liefern, so ist OnlineOCR trotzdem einen Versuch wert.

### 3.5 Schlussbemerkungen

Mit anderen Suchbegriffen wären sicher noch weitere Konverter gefunden worden, von denen einige möglicherweise noch bessere Resultate als die getesteten geliefert hätten. Aber im Alltag sind wir wohl höchst selten mit dem Problem konfrontiert, aus einem nur auf Papier oder als Foto vorliegenden Dokument ein Word- oder Excel-Dokument erzeugen zu müssen. Falls das doch mal vorkommen sollte, kann dieses Merkblatt weiterhelfen. Wer eine grössere Anzahl, oder immer wieder Dokumente konvertieren muss, ist mit einem auf dem eigenen Computer installierten Konversions-Programm besser beraten.