

Daten in Excel bereinigen

Datenquellen sauber importieren	1
Textimport-Assistent:	1
Alle Spalten und Reihen in Excel anzeigen	3
Spalten Transformieren	3
Duplikate entfernen	4

Das Informationsblatt listet die vier gängigsten Schritte, mit denen Daten in Excel bereinigt werden können. Bevor man allerdings mit der Bereinigung von Daten beginnt, ist es immer gut, eine Kopie der Originaldatei anzulegen.

Datenquellen sauber importieren

Excel hat häufig Probleme damit, Daten aus anderen Quellen richtig zu importieren. Beispielsweise werden Umlaute nicht richtig erkannt und können nicht korrekt dargestellt werden. Über den Menüpunkt **Daten** können spezielle Einstellungen für den Import von Textdateien (z.B. csv.), HTML-Dateien sowie Importe aus Datenbanken vorgenommen werden. Da sehr häufig Textdateien importiert werden, wird dieser Schritt genauer erklärt.



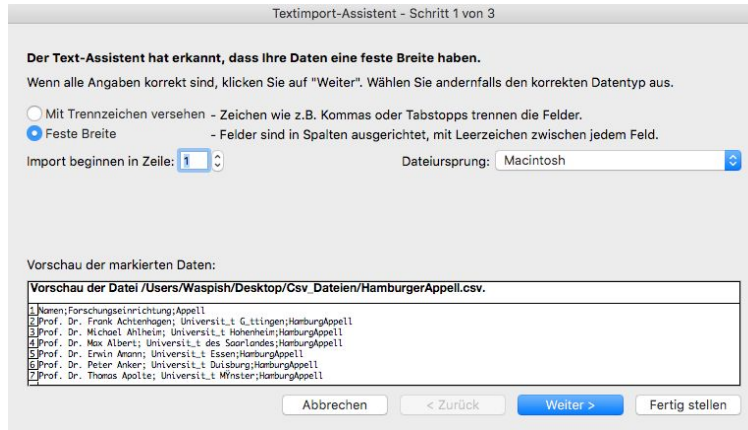
Der Menüpunkt Daten

Textimport-Assistent

Klickt man auf den Menüpunkt "Aus Text", öffnet sich der **Text-Assistent**. In drei Schritten können hier spezielle Einstellungen für den Import von Textdateien vorgenommen werden.

Schritt 1: Das Ursprungsformat festlegen

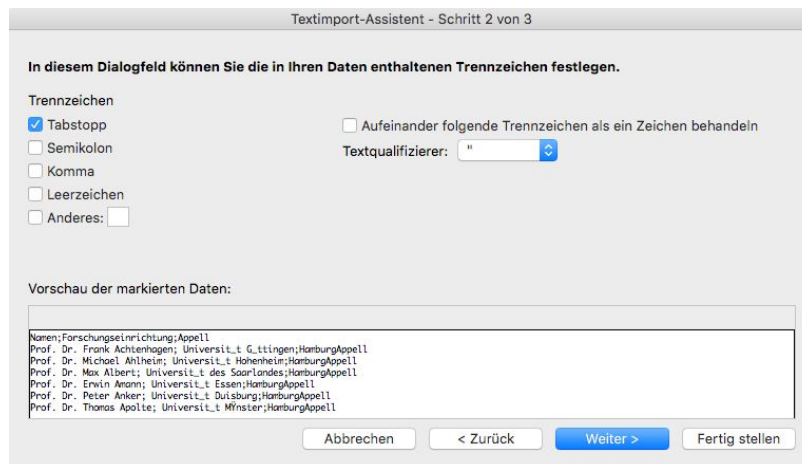
Im ersten Schritt kann der **Dateiursprung** der Import-Datei angegeben werden¹. Dieser muss allerdings längst nicht immer angepasst werden. In diesem Schritt kann auch angegeben werden, ab welcher Zeile in einer Datei Elemente importiert werden sollen.



Schritt 1: Das Ursprungsformat festlegen

Schritt 2: Das richtige Trennzeichen auswählen

Die einzelnen Felder eines Datensatzes sind durch ein festgelegtes Zeichen voneinander abgetrennt. Ein Trennzeichen. Klickt man auf den Punkt *“Mit Trennzeichen versehen”*, kann das richtige Trennzeichen eines Datensatzes angegeben werden, sodass Dateien sauber importiert werden können. Häufige Trennzeichen sind **Kommata**, **Semikola** oder **Tab**. **Leerzeichen** sind seltener, werden aber auch teilweise genutzt. In dem Feld **Anderes:** können andere Zeichen eingegeben werden.

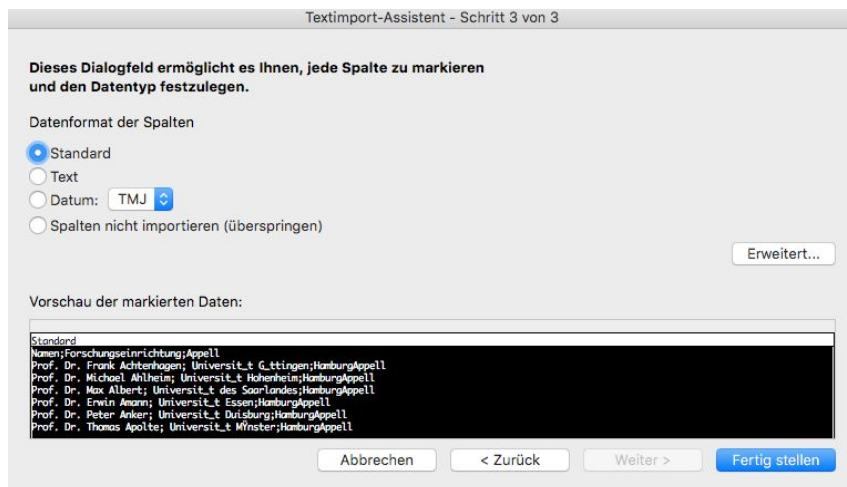


Schritt 2: Das richtige Trennzeichen auswählen

¹ Der Dateiursprung legt fest, welcher Zeichensatz in einem Datensatz verwendet wurde. Daten können in vielen verschiedenen Zeichensätzen dargestellt werden. Gängige Formate sind Unicode (UTF-8), ASCII, Macintosh, Latin-1 (ISO 8859-1).

Schritt 3: Das richtige Format für einzelne Spalte festlegen

Im dritten Schritt kann für jede zu importierende Spalte das richtige **Format** festgelegt werden. Beispielsweise kann unterschieden werden, ob es sich um ein Datum, eine Zahl oder eine Spalte handelt, in der Textelemente enthalten sind.



Schritt 3: Das richtige Format für einzelne Spalte festlegen

Alle Spalten und Reihen in Excel anzeigen

In Excel lassen sich Reihen und Spalten ausblenden. Bei Datensätzen mit vielen Einträgen kann dies Übersicht verschaffen. Für die Datenbereinigung sollten allerdings zu Beginn alle Reihen und Spalten eingebledet werden. Dazu oben links zwischen der "Spalte A" und der "Reihe 1" auf das Dreieck klicken (siehe Bild), um alle Zeilen und Spalten zu markieren. Per Rechtsklick lässt sich die Aktion "Einblenden" auswählen, sodass automatisch alle Spalten und Reihen angezeigt werden.



Alle Spalten & Zeilen auswählen

Spalten Transformieren

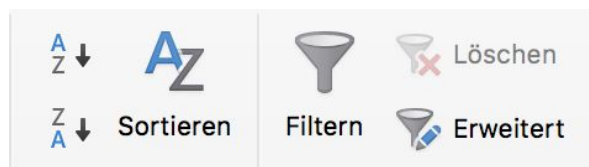
Damit Werte in einem Datensatz verglichen werden können, müssen sie einheitlich strukturiert vorliegen. Leerzeichen, unterschiedliche Schreibweisen und Formatformatierungen sorgen häufig für eine Unstrukturiertheit in den Daten. In der unten stehenden Tabelle sind fünf nützliche Excel-Funktionen² gelistet, mit denen Werte bereinigt und vereinheitlicht werden können.

² Funktionen sind vorgefertigte Berechnungen die in Excel integriert sind. Funktionen können auf zwei verschiedene Arten in Excel aufgerufen werden. (1) Über den Menüpunkt **Formeln** kann nach Funktionen gesucht werden. Der **Formel Generator** hilft einem bei der richtigen Anwendung der Funktion. (2) Indem man in eine Zelle ein Leerzeichen schreibt kommt man automatisch in den Funktions-Modus und kann die gewünschte Excel-Funktion eingeben und anwenden.

Funktionen	Beschreibung
=SÄUBERN()	<i>krusiv-</i> oder <u>unterstrich</u> -Formatierungen werden in einer Zelle entfernt.
=GLÄTTEN()	Leerzeichen zu Beginn und am Ende einer Zelle werden entfernt.
=GROSS1()	Alle Buchstaben in einer Zelle werden groß geschrieben.
=GROSS2()	Alle Anfangsbuchstaben eines Wortes in einer Zelle werden groß geschrieben.
=KLEIN()	Alle Buchstaben in einer Zelle werden klein geschrieben.

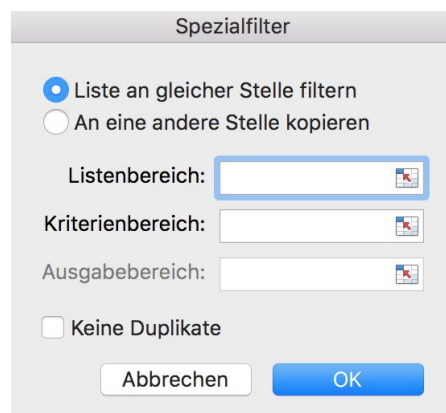
Duplikate entfernen

Häufig soll ein Datensatz keine doppelten Einträge enthalten und nur aus Unikaten bestehen. Um Duplikate herauszufiltern, gibt es eine eigene Funktion in Excel. Für diesen Schritt ist es besonders wichtig, dass die Daten einheitlich strukturiert vorliegen, damit Werte auch als Duplikate erkannt werden können. Hierfür in dem Menüpunkt **Daten** die Aktion **Filtern** auswählen.



Der Aktionspunkt Filtern

Anschließend auf **Erweitert** klicken, sodass sich der **Spezialfilter** öffnet. In dem Eingabefeld **Listenbereich** kann nun der gewünschte Bereich angegeben werden, in dem doppelte Einträge gefiltert werden sollen. Dafür auch bei "**keine Duplikate**" einen Haken setzen. Allerdings werden die doppelten Einträge standardmäßig nur ausgeblendet. Möchte man Duplikate aus einer Liste endgültig entfernen, kann die Option "**An eine andere Stelle kopieren**" ausgewählt werden. Anschließend wird eine neue Liste erstellt, die nur aus Unikaten besteht.



Der Spezialfilter