

nungen verwendet werden.

3.5 Tabellen

M. Trognitz

Tabellen werden verwendet, um Informationen strukturiert in Zellen zu speichern, die in Zeilen und Spalten angeordnet sind. Im Gegensatz zu analogen Tabellen bieten digitale Tabellenkalkulationen, die mit entsprechenden Programmen erstellt werden, eine Vielzahl an weiteren Funktionalitäten. Beispielsweise können Werte dynamisch mittels Formeln errechnet oder Grafiken aus den Daten erzeugt werden. Um solche und andere Funktionen zu erhalten, erfordert die Speicherung besondere Aufmerksamkeit.

Langzeitformate Für einfache Tabellen ohne interaktive Elemente wie Formeln oder für Tabellenkalkulationen bei denen es ausreicht die Ergebnisse der Formeln zu speichern, wird ein textbasiertes Format mit Trennzeichen für die Archivierung empfohlen.

Dazu eignet sich beispielsweise das CSV-Format, wobei als Trennzeichen ein Komma (,) und als Textbegrenzungszeichen das Anführungszeichen (") verwendet werden sollte, um den Vorgaben von RFC 4180 gerecht zu werden. Andere Trennzeichen und Textbegrenzungszeichen können in begründeten Ausnahmefällen ebenfalls eingesetzt werden und müssen entsprechend dokumentiert werden. Als Zeichenkodierung sollte UTF-8 ohne BOM verwendet werden.

Ein alternatives textbasiertes Format für Tabellen ist das TSV-Format, das als Trennzeichen das Tabulator-Zeichen (U+0009) verwendet. Auch TSV-Dateien sollten UTF-8 ohne BOM als Zeichenkodierung verwenden.

Bei Tabellenkalkulationen mit mehr als einem Arbeitsblatt, die als CSV- oder TSV-Datei gespeichert werden sollen, muss jedes Arbeitsblatt gesondert gespeichert werden. Der Name des Arbeitsblattes wird dabei an den Tabellennamen im Dateinamen am besten durch ein Unterstrich (_) getrennt angefügt (z. B. Tabellenname_Blatt1.csv, Tabellenname_Blatt2.csv usw.).

Tabellenkalkulationen, deren zusätzlichen Funktionalitäten erhalten bleiben sollen, werden am besten in einem offenen auf XML basierenden Format gespeichert, wie beispielsweise XLSX oder ODS. Ersteres ist das Standardformat, das in Microsoft Excel seit 2007 verwendet wird und auch von Microsoft entwickelt wurde. Letzteres ist das Format für Tabellenkalkulationen, welches in OpenOffice oder LibreOffice verwendet wird. ODS ist ein Teil vom OpenDocument Format (ODF) und wurde von einem technischen Komitee unter der Leitung der *Organization for the Advancement of Structured Information Standards* (OASIS) entwickelt.

Grafiken, die in Tabellenkalkulationen anhand der Daten erstellt wurden, müssen zusätzlich exportiert und gesondert in einem geeigneten Format gespeichert werden.

Tabellen können auch im XML-Format gespeichert werden. Es gibt eine ganze Reihe an DTDs oder XSDs, die hier als Grundlage dienen können, wie beispielsweise ISO 12083, das Tabellenmodell von CALS oder das Schema der TEI. Auch das Speichern von Tabellen im HTML-Format ist möglich. In jedem Fall muss die zugrundeliegende DTD oder XSD angegeben und gegebenenfalls

archiviert werden. Die Dateien sollten UTF-8 ohne BOM als Zeichenkodierung verwenden.

Wenn neben den eigentlichen Daten in den einzelnen Zellen auch das Aussehen der Tabelle archiviert werden soll, kann neben einer CSV-, TSV-, XLSX- oder ODS-Datei zusätzlich eine Kopie der Tabelle als PDF/A gespeichert werden. Bei der Erstellung von Tabellen sollte aber darauf geachtet werden, dass Informationen nicht nur durch Formatierungsangaben, wie beispielsweise die Farbe von Zellen, vermittelt werden.

Hinweis: Eingebettete Bilder oder andere Medien sollten zusätzlich separat gespeichert werden. Außerdem muss beachtet werden, dass Links oder dynamische Inhalte, nicht immer dauerhaft erhalten bleiben.

Format	Begründung
✓ CSV	Das textbasierte CSV-Format sollte mit einem Komma als Trennzeichen und mit Anführungszeichen als Textbegrenzungszeichen verwendet werden. Ausnahmen müssen dokumentiert werden. Die Zeichen sollten in UTF-8 ohne BOM kodiert sein.
TSV	TSV (MIME-Type text/tab-separated-values) ist ein textbasiertes Format, welches das Tabulator-Zeichen (U+0009) als Trennzeichen verwendet. Die Zeichen sollten in UTF-8 ohne BOM kodiert sein.
ODS	ODS basiert auf XML und ist Teil vom Open-Document Format. Eingebettete Bilder und Medien müssen gesondert gespeichert werden.
XLSX	XLSX ist das auf XML basierende Format von Microsoft. Eingebettete Bilder und Medien müssen gesondert gespeichert werden.
XML, HTML	Tabellen im textbasierten XML- oder HTML-Format können ebenfalls archiviert werden. XML-Dateien benötigen zusätzlich eine DTD-Datei oder das XML Schema. Die Zeichen sollten in UTF-8 ohne BOM kodiert sein.
~ PDF/A	Wenn neben der Daten auch das Aussehen der Tabelle erhalten bleiben soll, eignet sich PDF/A am besten. Zusätzlich sollten die tabellarischen Daten in einem zugänglichen Format, wie etwa CSV, gespeichert werden.
SXC	SXC ist ein Vorgängerformat von ODS, weshalb letzteres auch bevorzugt werden sollte.
✗ XLS	Das XLS-Format von Microsoft eignet sich nicht zur Archivierung, da es proprietär ist und die Inhalte nicht textbasiert gespeichert werden.

Dokumentation Neben den allgemeinen Angaben zu Einzeldateien, wie sie in dem Abschnitt Metadaten in der Anwendung ab Seite 37 gelistet sind, werden für Tabellen und Tabellenkalkulationen weitere Angaben benötigt.

Um die Verständlichkeit einer Tabelle auch für Dritte zu gewährleisten, müssen Name und Zweck der jeweiligen Tabelle und der einzelnen Arbeitsblätter bekannt sein. Jede Spalte benötigt eine Überschrift, und zusätzlich müssen die verwendeten Formatvorgaben, Abkürzungen, Codes, Wertelisten und sonstige Terminologien dokumentiert werden. Wenn Maßeinheiten nicht direkt aus der Tabelle ersichtlich sind, müssen diese ebenfalls gesondert dokumentiert werden.

Um sicher zu gehen, dass die Tabelle auch vollständig vorliegt, sollten die Anzahl der Spalten, Zeilen und gegebenenfalls auch der Arbeitsblätter angegeben werden.

Tabellen in textbasierten Formaten brauchen Angaben zu den verwendeten Trennzeichen, Textbegrenzungszeichen und der Zeichenkodierung.

Für Tabellenkalkulationen müssen bei Bedarf weitere Informationen zu Relationen, Formeln und Makros dokumentiert werden. Eingebettete Medien, wie etwa Bilder, sollten separat gespeichert und archiviert werden und in einer Liste weiterer Dateien aufgeführt werden. Dies gilt ebenfalls für Grafiken, die aus den Daten in der Tabellenkalkulation erzeugt wurden.

Die hier angegebenen Metadaten sind als minimale Angabe zu betrachten und ergänzen die angegebenen Metadaten für Projekte und Einzeldateien in dem Abschnitt Metadaten in der Anwendung ab Seite 37.

Metadatum	Beschreibung
Beschreibung der Tabelle oder des Arbeitsblattes	Welchen Zweck verfolgt die Tabelle oder das Arbeitsblatt?
Namen der Arbeitsblätter	Titel der Arbeitsblätter.
Spaltenüberschrift	Jede Spalte einer Tabelle muss einen Namen haben.
Spaltenbeschreibung	Beschreibung und Auflistung der in der jeweiligen Spalte verwendeten Formatvorgaben, Abkürzungen, Codes, Wertelisten oder Fachvokabulare. Welche Maßeinheiten liegen vor?
Anzahl Zeilen	Wie viele Zeilen enthält die Tabelle?
Anzahl Spalten	Wie viele Spalten enthält die Tabelle?
Anzahl Arbeitsblätter	Anzahl der Arbeitsblätter in einer Tabellenkalkulation.
Trennzeichen	Angabe des verwendeten Trennzeichens bei textbasierten Speicherformaten, wie CSV.
Textbegrenzungszeichen	Angabe des verwendeten Textbegrenzungszeichens bei textbasierten Speicherformaten, wie CSV.
Zeichenkodierung	Angabe der verwendeten Zeichenkodierung bei textbasierten Speicherformaten, wie CSV oder TSV.
Relationen	Welche Relationen gibt es innerhalb der Tabellenkalkulation?
Formeln	Welche Formeln werden in der Tabellenkalkulation verwendet?
Makros	Welche Makros gibt es in der Tabellenkalkulation?
Aus Daten erzeugte Grafiken	Aus den Daten erzeugte Grafiken müssen zusätzlich separat gespeichert werden und in die Liste der weiteren Dateien aufgenommen werden.
Sprache	In welchen Sprachen ist das Dokument verfasst? Sprachkennungen nach ISO 639 angeben.
Identifikator	Wenn das Dokument bereits veröffentlicht wurde und eine ISBN oder einen anderen persistenten Identifikator erhalten hat, sollte dieser angegeben werden.
weitere Dateien	Liste von eingebetteten Medien, wie Bilder, die zusätzlich separat gespeichert wurden. Liegt eine Dokumentationsdatei für das Dokument vor, muss diese ebenfalls genannt werden.

Weitere Metadaten sind methodenabhängig und können in den jeweiligen

Abschnitten nachgelesen werden.

Vertiefung

Tabellen bestehen in ihrer einfachsten Form aus Zeilen und Spalten, deren Überschneidung eine Zelle bilden, in der ein Wert eingetragen wird. Die Namen der Spalten werden in die erste Zeile, der Kopfzeile (*header*) eingetragen und in der ersten Spalte, der Vorspalte, können die Zeilenbezeichnungen stehen.

Digitale Tabellen und Tabellenkalkulationen (*spreadsheets*) bieten erweiterte Funktionalitäten.

Für die Langzeitarchivierung von Tabellen sind strukturierte Textdateien mit Trennzeichen (*delimiter separated values*), wie CSV- oder TSV-Dateien, gut geeignet. Um Tabellen als XML-Dateien zu speichern, gibt es verschiedene XML Schemata und Dokumenttypdefinitionen.

	A	B	C	D
1	A1	B1	C1	D1
2	A2	B2	C2	D2
3	A3	B3	C3	D3
4	A4	B4	C4	D4

Abb. 3.11: In der Tabelle ist die Spalte *B* blau markiert. Die Zeile 3 ist gelb markiert. Die grün hervorgehobene Überschneidung aus der Zeile und der Spalte ist die Zelle *B3*.

Tabellenkalkulationen Als Tabellenkalkulation wird sowohl das Programm, als auch die damit erstellte Tabelle oder der Satz an Tabellen bezeichnet. Im Folgenden ist die resultierende Datei gemeint.

Eine Tabellenkalkulation kann mehrere Tabellen, sogenannte Arbeitsblätter (*worksheets* oder *spreadsheets*) enthalten. Die Daten selbst können innerhalb der Tabellenkalkulation verwendet werden, um mittels Formeln neue Daten zu erzeugen oder Grafiken zu erstellen. Dazu erlauben die Programme auch eine Referenzierung auf Werte in anderen Zellen, die sich sogar in einem anderen Arbeitsblatt befinden können. Es handelt sich dabei um Relationen.

Die gängigsten Tabellenkalkulationsprogramme ermöglichen die Erstellung und Verwendung von Makros, die es ermöglichen typische Befehlsfolgen und Bedienungsschritte aufzuzeichnen und automatisiert zu wiederholen.

Formatierungsangaben können in digitalen Tabellen und Tabellenkalkulationen ebenfalls gemacht werden. Dabei können wie in Textdokumenten nicht nur die Schriftart, Schriftgröße und ähnliches angepasst werden, sondern auch das Aussehen von Zellen, Zeilen und Spalten mittels Angaben zu Rahmenlinien und Hintergrundfarben. Soll der durch die Formatierungsangaben erzeugte optische Eindruck ebenfalls archiviert werden, empfiehlt sich eine zusätzliche Speicherung der Tabellenkalkulation als PDF-Dokument.

Strukturierte Textdateien mit Trennzeichen Tabellen können in einem textbasierten Format gespeichert werden. Dabei handelt es sich um eine Textdatei, die auf eine bestimmte Weise strukturiert ist. Für Tabellen gibt es insbesondere die Formate CSV (*comma-separated values*) und TSV (*tab-separated values*), die sich nur geringfügig voneinander unterscheiden.

Jede Zeile einer CSV- oder TSV-Datei entspricht der Zeile einer Tabelle. Die einzelnen Zellen werden durch sogenannte Trennzeichen voneinander getrennt.