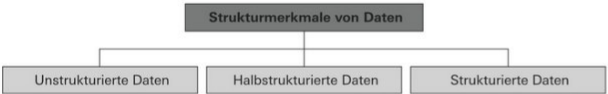


## 2 Strukturmerkmale von Daten

Um Daten auswerten zu können, sind Strukturmerkmale von entscheidender Bedeutung. Dabei kann zwischen unstrukturierten, halbstrukturierten und strukturierten Daten unterschieden werden:

[2-1] Strukturmerkmale von Daten



### 2.1 Unstrukturierte Daten

Unstrukturierte Daten sind digitalisierte Daten, die in einer nicht formalisierten Struktur vorliegen. Das heisst, dass keine erkennbare Struktur oder ein einheitlicher Aufbau vorliegt, der eine automatische Bearbeitung zulässt. Dazu gehören etwa digitalisierte Tonaufnahmen oder Fliesstexte ohne Gliederung.

#### Beispiel

Früher mussten die Kunden des Fitnesscenters «Fusskick» ihre Mitgliedschaft schriftlich per Brief kündigen. Dabei hat jeder Kunde seine Kündigung individuell erstellt. Manche Kunden haben auch handschriftliche Briefe eingereicht. Sämtliche Briefe wurden von Isabelle Baumann eingescannt (digitalisiert) und archiviert.

Aufgrund der «individuellen Struktur» dieser Briefe haben Sie es mit unstrukturierten Daten zu tun. Eine automatisierte Auswertung ist nicht möglich. So kann Fabian Moser die Kündigungen z. B. nicht nach Kündigungsgründen auswerten. Die Briefe müssen einzeln gelesen und (weiter-)bearbeitet werden. Für unstrukturierte Daten gilt also, dass diese zuerst in eine formale Struktur gebracht werden müssen, bevor sie ausgewertet werden können.

### 2.2 Halbstrukturierte Daten

Halbstrukturierte Daten sind digitalisierte Daten, die keiner vorgegebenen Struktur folgen, aber bestimmte Strukturinformation beinhalten. Solche Daten tragen also Informationen über den Aufbau in sich und werden auch schwach- oder semistrukturierte Daten genannt.

#### Beispiel

Isa Bauman benutzte viele Jahre lang Karteikarten für die Kundenverwaltung. Diese Karten wurden in der Zwischenzeit eingescannt (digitalisiert).

[2-2] Karteikarte

B	
Kundennummer:	33
Anrede:	Herr
Name:	Balmelli
Vorname:	Marco
Strasse:	Schönstr. 12
Ort:	8000 Zürich
Eintritt:	01.01.2005
Abonnenten-Typ:	Student
Gebühr:	500.00 Fr.

Auf den digitalisierten Karteikarten ist eine gewisse Struktur erkennbar, d. h., es werden Strukturinformationen mitgeliefert. So ist z. B. auf einem Blick erkennbar, dass es sich beim Kunden mit der Nummer 33 um Marco Balmelli handelt, der ein Studenten-Jahresabo gelöst hat.

Weit verbreitet sind halbstrukturierten Daten in Form von **XML**<sup>[1]</sup>. Diese Auszeichnungssprache dient als Format für den Datenaustausch im Internet und wird zudem von vielen Applikationen als Format für die Datenablage verwendet. Auch bei den halbstrukturierten Daten gilt, dass sie zuerst in eine formale Struktur gebracht werden müssen, bevor eine Auswertung stattfinden kann.

### 2.3 Strukturierte Daten

Strukturierte Daten weisen eine gleichartige Struktur auf, d. h., der Aufbau der Daten ist vorgegeben, einheitlich und klar erkennbar. Strukturierte Daten setzen eine eindeutige Festlegung voraus, wie und wo die Daten gespeichert werden. Dies ist beispielsweise bei einem Datenbanksystem<sup>[2]</sup> der Fall. Um Auswertungen erstellen zu können, müssen strukturierte Daten vorliegen. Falls die Daten in unstrukturierter oder halbstrukturierter Form vorliegen, müssen sie zuerst in eine strukturierte Form gebracht werden.

Digitalisierte Daten können bezüglich der Strukturmerkmale wie folgt gegliedert werden:

```
graph TD; A[Strukturmerkmale von Daten] --> B[Unstrukturierte Daten]; A --> C[Halbstrukturierte Daten]; A --> D[Strukturierte Daten];
```

**Unstrukturierte** Daten verfügen über keinerlei Struktur. **Halbstrukturierte** Daten sind Daten, die keiner allgemeinen Struktur unterliegen, sondern einen Teil der Strukturinformation mit sich tragen. Für Auswertungen sind **strukturierte Daten** notwendig, wie sie z. B. in einer Datenbank vorliegen.

[1] Vergleichen Sie dazu Kapitel 6.3.3, S. 35.

[2] Vergleichen Sie dazu Kapitel 4.1, S. 24.

Repetitionsfragen	
4	Sie erhalten eine E-Mail. In welcher Struktur liegt diese vor? Begründen Sie Ihre Antwort.
5	Weshalb können unstrukturierte Daten schlecht ausgewertet werden?
6	Welche Daten beinhalten Strukturinformationen?

## 3 Dateiatien und ihre Auswertbarkeit

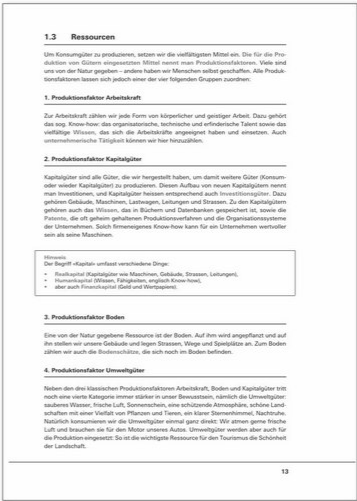
Im Kapitel 2.3, S. 15 haben Sie erfahren, dass strukturierte Daten am besten und einfachsten ausgewertet werden können. In diesem Kapitel werden folgende **Dateiatien** auf ihre Auswertungsmöglichkeit hin untersucht:

- Textdatei
- Präsentation
- Tabelle
- Diagramm
- Grafikdatei
- Audiodatei
- Videodatei
- Ausführbares Programm

### 3.1 Textdatei

Bei der Textdatei handelt es sich um eine Dateiart, die für die Verwaltung von reinem Text verwendet wird. Solche Dateien sind sehr flexibel, wiederverwendbar und werden z. B. für Korrespondenzen, Geschäftsberichte, Protokolle und HTML-Dateien oder für den Source-code von Programmen verwendet. Hier ein Beispiel dazu:

[3-1]      Ausschnitt aus einer Textdatei



Im Allgemeinen kann zwischen **unformatierten** und **formatierten Textdateien** unterschieden werden, wobei unformatierte Textdateien auch einfache Textdateien genannt werden. Sie lassen sich durch folgende Eigenschaften und Beispiele charakterisieren:

	Unformatierte Textdateien	Formatierte Textdateien
<b>Merkmale</b>	<ul style="list-style-type: none"> <li>Enthalten ausschliesslich Texte</li> <li>Können mit jedem Editor betrachtet und bearbeitet werden</li> </ul>	<ul style="list-style-type: none"> <li>Enthalten sowohl reine Texte als auch Informationen über dessen Formatierung (z.B. Überschriften, Tabellen)</li> <li>Setzen ein Anwendungsprogramm voraus, um den Text korrekt darzustellen</li> </ul>
<b>Beispiele</b>	<ul style="list-style-type: none"> <li>Notepad</li> <li>Wordpad</li> </ul>	<ul style="list-style-type: none"> <li>MS Word</li> <li>Lotus Word Pro</li> </ul>

Textdateien können auf folgende **Arten** entstehen:

- **Manuelle Eingabe:** Die wohl bekannteste Möglichkeit, Textdateien zu erstellen, geschieht über Mensch und Tastatur. Sie können jede Information zusammentragen und über die Tastatur erfassen.
- **Scannen und Texterkennung:** Doch was machen Sie, wenn Sie ein gedrucktes Buch elektronisch erfassen möchten? Die manuelle Eingabe von Tausenden von Seiten wäre zu aufwendig. Sie haben nun die Möglichkeit, mit einem Scanner sowie einer Texterkennungssoftware bereits gedruckte Texte wieder zu Textdateien umzuwandeln. Dabei liest der Scanner Seite um Seite ein und ein spezielles Programm erkennt Muster und ordnet diesen Buchstaben zu.
- **Export und Import:** Applikationen und Programme können Textdateien erzeugen, beispielsweise um Daten an eine andere Applikation zu liefern, wenn die Quelldaten nicht kompatibel sind oder ein Zugriff nicht möglich ist. Dieser Austausch mittels sogenannten Standardformaten wird in Kap. 6.3.3, S. 35 kurz vorgestellt.

Die **Entstehungsart einer Textdatei** wirkt sich massgeblich auf die Auswertbarkeit der Inhalte bzw. Daten aus. Von Hand geschriebene Briefe lassen sich z. B. nur schlecht auswerten, während hoch strukturierte Texte in standardisierten Formaten gut auswertbare Inhalte bzw. Daten aufweisen und sich auch gut für den Datenaustausch eignen.

## 3.2 Präsentation

Präsentationen bestehen meist aus einer Kombination von Grafiken, Bildern und Texten, die in Form von **Folien** in einer bestimmten Reihenfolge abgespeichert werden. Eine Präsentation wird in der Regel von Hand mithilfe von vorgefertigten Elementen (z. B. Cliparts) zusammengestellt. Masterfolien, die das Layout für alle Folien bestimmen, vereinfachen die Arbeit und gewährleisten ein einheitliches Erscheinungsbild. Präsentationen werden als Binärdatei<sup>[1]</sup> abgespeichert und können daher nicht mit einem regulären Editor gelesen werden.

Manchmal werden Präsentationen mit Ton- und Videosequenzen angereichert und können als Show mit audio-visuellen Effekten abgespielt werden. Solche **multimedialen Präsentationen** lassen sich nur mit einer speziellen Präsentationssoftware herstellen, abspeichern und abspielen.



Eine Präsentation kann entweder als **unterstützendes Medium** im Rahmen eines Vortrags (z. B. im Unterricht) oder als **selbstablaufende Vorführung** (die z. B. im Wartezimmer oder im Schaufenster abläuft) eingesetzt werden.

Multimediale Präsentationen lassen sich nur schlecht auswerten, da keine einheitliche Struktur vorliegt. Eine Volltextsuche über eine oder mehrere Folien hinweg ist aber meist möglich.

## 3.3 Tabelle

Eine Tabelle besteht aus Spalten und Zeilen, die **gleichartige Daten** zusammenfassen. Meist stehen in den Zeilen verschiedene Objekte, während in den entsprechenden Spalten die zugehörigen Eigenschaften aufgeführt werden. Hier ein Ausschnitt einer Tabelle mit zwei Zeilen (Abonnenten des Fitnesscenters «Fusskick») und zwei Spalten (Kundennummer und Name):

[3-4] Tabelle (Beispiel)

Kundennummer	Name
11	Balmelli
22	Bürgin

Tabellen werden also zur Speicherung von Daten mit gleichartiger Struktur verwendet und können in Tabellenkalkulationsprogrammen (z. B. MS Excel) verwaltet werden. Aber auch Datenbankprogramme führen mehrere Tabellen, die miteinander verknüpft sind.<sup>[1]</sup>

[1] Datei, die aus einer Abfolge von 8-Bit-Daten oder aus einem ausführbaren Code besteht und sich dadurch von einer Datei unterscheidet, die lesbaren Text im ASCII-Format enthält.

[1] Vergleichen Sie dazu Kap. 5, S. 26.

Die Daten einer Tabelle werden meist über eine vorstrukturierte Eingabemaske erfasst oder bearbeitet. Tabellen sind also stark strukturiert und lassen sich daher gut auswerten.

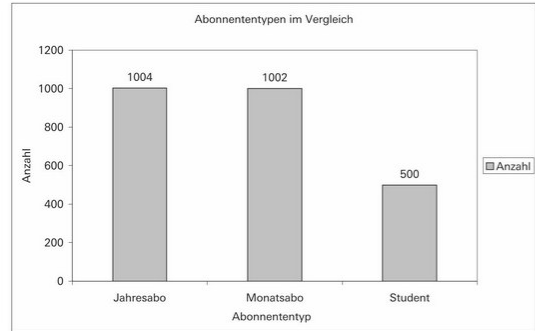
**Beispiel**

Interessierte Personen können direkt über die Webseite des Fitnesscenters «Fusskick» ein Abonnement lösen. Die Kunden- und Anmeldedaten werden in einem Standardformular abgefragt und direkt in eine Tabelle eingegeben.

### 3.4 Diagramm

Ein Diagramm ist eine **grafische Darstellung** von Zahlen bzw. Werten und anderen Daten oder Informationen. Grafisch aufbereitete Fakten und Zusammenhänge lassen sich leichter verstehen und vermitteln als beispielsweise verbale Beschreibungen oder Tabellen mit umfangreichem Zahlenmaterial. Hier ein Diagramm, dass die Mitglieder des Fitnesscenters «Fusskick» gegliedert nach Abonententypen darstellt:

[3-5] Diagramm (Beispiel)



Ein Diagramm ist meistens das Produkt bereits aufbereiteter Daten bzw. das Ergebnis bereits ausgewerteter Daten. Das bedeutet, dass die Basisdaten vorher erfasst oder aus anderen Systemen importiert worden sind. Oft werden in einem Diagramm nicht einzelne Datensätze, sondern bereits berechnete Werte angezeigt. Ein Diagramm kann daher nicht mehr elektronisch ausgewertet werden. Wenn weitere Auswertungen verlangt werden, muss man auf die **Quellendaten** zurückgreifen.

### 3.5 Grafikdatei

Eine Grafik ist eine Abbildung (Piktogramm, Zeichnung, Fotografie, Illustration), die digital gespeichert wurde. Ein Text, der eingescannt und als Abbildung gespeichert wurde, kann ebenfalls wie eine Grafikdatei behandelt werden. Hier ein Bild, das Blüten einer Orchidee zeigt und als Grafikdatei bearbeitet werden kann:

[3-6] Grafikdatei (Beispiel)



Grafikdateien entstehen bei der digitalen Fotografie, beim Einscannen von Texten und/oder Bildern sowie beim Zeichnen mit einem Zeichnungsprogramm. Eine Grafik ist eine reine Binärdatei und somit unstrukturiert und schlecht auswertbar. Je nach Entstehungsart kann eine Grafik in verschiedenen Formaten gespeichert werden. Dabei wird zwischen Vektor- und Rastergrafiken unterschieden. Weitverbreitet sind die Formate BMP, EPS, GIF, JPG, PNG und TIF.

### 3.6 Audiodatei

In einer Audiodatei können Tondokumente jeder Art (Musik, Sprache, Geräusche etc.) verwaltet werden. Während Tonaufnahmen früher auf analogen Vinyl-Schallplatten oder Magnetbändern aufgezeichnet wurden, werden sie heute meist digital gespeichert und bearbeitet. Danach kann eine Audiodatei entweder auf CD oder DVD gebrannt und verkauft oder über einen Online-Shop im Internet vertrieben werden.

Audiodateien entstehen beispielsweise bei Radiosendungen im Internet. Dabei wird das gesendete Programm gespeichert und gleichzeitig zum Empfänger übertragen. Diese Technik heisst Streaming. Beim Streaming muss die Übertragungsrate höher sein als die Abspielgeschwindigkeit beim Empfänger, da sonst keine Kontinuität gewährleistet werden kann. Je nach Übertragungsleistung werden die Audiodaten mehr oder weniger stark komprimiert. Dies hat zur Folge, dass die Qualität bei tiefen Übertragungsraten entsprechend schlechter wird. Um Schwankungen bei der Übertragungskapazität auszugleichen, wird auf der Empfängerseite ein Puffer angelegt. Aus diesem Grund erfolgt die Wiedergabe leicht verzögert (ca. 30 Sekunden).

Die bekanntesten Audiodateien sind wohl Musikstücke auf CDs oder im MP3-Format. Akustische Warn- und Signaltöne im Computer kommen von Audiodateien, genauso wie Sprachausgaben und Hintergrundmusik auf Webseiten.

Wie Grafikdateien sind auch Audiodateien reine Binärdateien und lassen sich daher kaum auswerten. Zwar sind begleitende Informationen wie z. B. der Interpret eines Titels als Text gespeichert, aber so in der Datei integriert, dass sie nur mit einem entsprechenden Audio-Player angezeigt werden können. Eine automatische Auswertung ist nicht möglich.



### 3.7 Videodatei

Ein Video besteht aus einer raschen Abfolge von Einzelbildern (ca. 25 Bilder pro Sekunde). Ein herkömmliches Fernsehsignal produziert ca. 166 Megabit pro Sekunde. Diese Datenmenge wird mit mehr oder weniger Verlust komprimiert. Je nach Kompressionsregeln wird entweder jedes Einzelbild komprimiert oder es werden mehrere Bildabfolgen miteinander verglichen und Redundanzen herausgefiltert. Die meisten Kameras speichern Filme digital. So entstehen sowohl im professionellen als auch im privaten Bereich zahlreiche Videodateien, die u. a. auch über das Internet angeboten werden (z. B. über die Plattform Youtube). Satelliten und Kabelfernsehen strahlen heute sowohl analoge als auch digitale Videosendungen aus. Aber auch Computerspiele basieren auf Videosequenzen, die durch den Spieler beeinflusst werden können. Wie die Grafik- und Audiodateien sind auch die Videodateien Binärdateien und deshalb nicht auswertbar.

### 3.8 Ausführbares Programm

Ausführbare Programme sind Programme, die gestartet werden können und dann selbstständig eine Funktion ausüben. Es gibt zwei Arten von Programmen, die der Computer ausführen kann:

- Bei sogenannten **Scriptsprachen** werden die vom Menschen eingegebenen Befehle bei jeder Ausführung durch einen Code-Interpreter interpretiert und ausgeführt. Dieser Programmcode ist als einfache Textdatei gespeichert.
- Bei anderen **Programmiersprachen** müssen die vom Menschen eingegebenen Befehle (Source-Code) mithilfe eines Compilers kompiliert, das heisst in Maschinensprache umgewandelt werden. Die Maschinensprache setzt sich nur aus 0 und 1 zusammen und kann vom Menschen nicht gelesen, vom Computer aber direkt ausgeführt werden. Programme in Maschinensprache sind daher viel effizienter als interpretierte Sprachen.

Der Programmcode wird durch den Programmierer als einfacher Text erfasst. Bei Interpreter-Sprachen ist dieser Programmcode gleichzeitig ein ausführbares Programm. Bei Compiler-Sprachen entstehen ausführbare Programme bei der Kompilation des Programmcodes. Im Internet gibt es JAVA-Applets und C++-Programme, die kompiliert sind. Im Gegensatz dazu sind JAVA-Script sowie PHP und ASP Interpreter-Sprachen. Bei Grossrechnern (Banken, Versicherungen) mit riesigen Datenvolumen werden für die Datenbearbeitung ausschliesslich kompilierte Programme eingesetzt, weil diese hohe Datenmengen rascher verarbeiten können.

Wie bei Textdateien kann auch der Programmcode von ausführbaren Programmen nur in beschränktem Mass ausgewertet werden. Sobald der Code aber einmal kompiliert worden ist und als ausführbares Programm vorliegt, kann er nicht mehr ausgewertet werden. Der bestehende Source-Code im Textformat kann jedoch weiterhin ausgewertet werden.

Ein Computer kann mit unterschiedlichen Arten von Dateien arbeiten. Jede Dateiart gehört zu einem speziellen Programm, das wiederum eine bestimmte Funktion ausführt, z. B. eine Präsentation.

In diesem Lehrmittel werden unterschiedliche Dateiarten für Dateien im Hinblick auf deren Auswertbarkeit aufgezeigt. Diese lassen sich wie folgt zusammenfassen:

Dateiart	Gut geeignet für Auswertungen	Geeignet für Auswertungen	Schlecht geeignet für Auswertungen
Textdatei	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Präsentation	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Tabelle	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Diagramm	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Grafik	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Audiodatei	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Videodatei	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Ausführbares Programm	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

### Repetitionsfragen

- |   |   |
|---|---|
| 7 | Sie möchten in einem Diagramm die Werte ändern und die Verkaufszahlen vom August hinzufügen. Wie müssen Sie vorgehen bzw. wo müssen Sie diese Änderung vornehmen? |
| 8 | Weshalb können Audiodateien nicht ausgewertet werden?   |
| 9 | Was ist ein Diagramm? Antworten Sie möglichst knapp und prägnant.   |