



# ZUSAMMENFASSUNG M101

Zusammenfassung zur Informatik-Prüfung über  
Modul 101.

Exposee

Zusammenfassung zur Informatik-Prüfung vom 03.07.2018 über das Modul 101 – Daten  
charakterisieren, aufbereiten und auswerten.

RaviAnand Mohabir  
ravianand.mohabir@stud.altekanti.ch  
<https://dan6erbond.github.io>

## Inhalt

1	Grundbegriffe .....	2
2	Strukturierung von Daten .....	2
3	Daten charakterisieren .....	1
3.1	Daten auswerten .....	1
3.2	Datenqualität einschätzen .....	1
4	Daten aufbereiten .....	1
4.1	Daten in die geeignete Form bringen .....	1
4.2	Daten zur Verfügung stellen .....	1
5	Daten auswerten .....	2
5.1	Filter bestimmen .....	2
5.1.1	Excel Spezialfilter .....	2
5.2	Abfragen mit Berechnung .....	2
6	Daten darstellen .....	1
6.1	Diagramme erzeugen und gestalten .....	1
6.1.1	Ergebnisse durch Darstellung gewichten .....	1
7	Relationale Datenbanken .....	1
7.1	Grundbegriffe .....	1
7.2	Probleme von normalen Tabellen .....	1
7.3	Relationale Datenbanken .....	2
7.3.1	Redundanzen entfernen .....	2
7.3.2	Tupel eindeutig identifizierbar machen .....	2
7.3.3	Beziehungen (Assoziationen) zwischen den Relationen herstellen .....	2

**Status:**      ☐ in Bearbeitung      ☒ Beendet



# 1 Grundbegriffe

## Nachrichten

Nachrichten sind Meldungen, welche von uns wahrgenommen werden und Informationen zwecks Weitergabe enthalten.

## Redundanz

Redundanzen sind Inhalt welche keinen unmittelbaren Neuigkeitswert enthalten und in der Datenbanktechnik unerwünscht sind, in der Nachrichtentechnik jedoch bewusst eingesetzt werden.

## Daten

Daten sind neutrale Elemente, welche eine erkennungsfähige Form haben. Sie sind entweder digital oder analog und können strukturiert sowie unstrukturiert sein. Daten sind maschinell verarbeitbar.

## Informationen

Informationen haben Inhalt und Bedeutungsgehalt. Sie enthalten keine irrelevanten oder redundanten Teile und sind kontextualisierte Daten. D.h. sie sind der Situation angepasst.

## Repräsentation

Durch Codierung wird die Information zur Speicherung in Form von Daten dargestellt.

## Abstraktion

Die Informationen werden aus codierten Daten durch Interpretation dieser zurückgewonnen.

## Wissen

Wissen entsteht, wenn Informationen mit anderen Informationen vernetzt werden und somit im Kontext eine andere Qualität erlangen oder zur Lösung von Problemen eingesetzt werden können.

# 2 Strukturierung von Daten

Daten repräsentieren Informationen in Form von Text, logischen Werten, Zahlen, Programmen, Bilder oder Musik. Sie kommen somit in verschiedenen Datentypen vor.

## Datentypen

Datentypen sind Zusammenfassungen von Objekten gleicher Ort und unterstützen definierte und zulässige Operationen. Bspw. Zahl.

## Datenstruktur

Zusammengehörige Daten werden als Struktur definiert. Diese unterstützen auch definierte und zulässige Operationen. Bspw. Datum.

Daten können unterschiedlich erfolgreich und effizient verarbeitet dargestellt werden. Dies hängt sehr stark davon ab, wie strukturiert die Rohdaten geliefert werden:

- Unstrukturiert: Fliesstext
- Schwach strukturiert / semistrukturiert: Tabelle
- Strukturiert: Datenbank



### 3 Daten charakterisieren

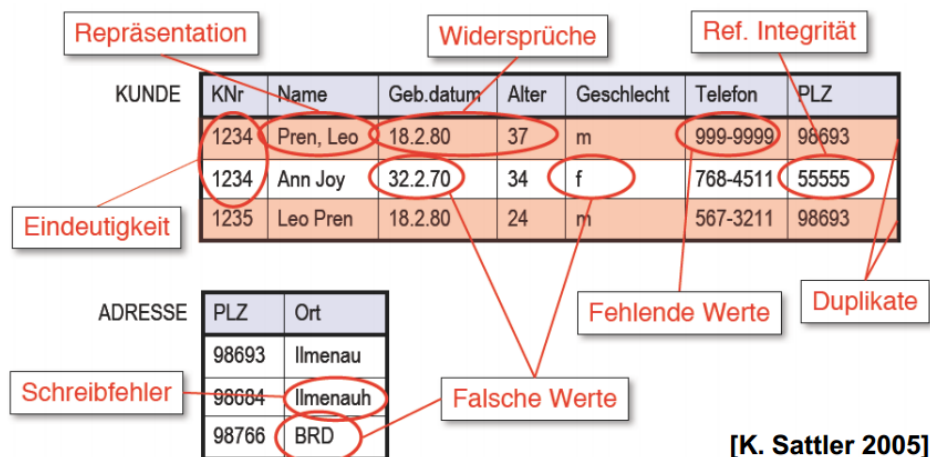
#### 3.1 Daten auswerten

Um die Auswertung zu beginnen, muss man die Auswertungswünsche kennen. Die Anforderungen sind oft nicht aus der IT-Sicht zu sehen. Man muss die Quellen bestimmen und eventuell mit internen/externen Datenquellen ergänzen.

#### 3.2 Datenqualität einschätzen

Die Datenqualität bezeichnet die Relevanz und Korrektheit von Informationen. Sie bestimmt die Eignung der Daten als Grundlage für die Weiterarbeit.

Es kann folgende Probleme mit der Datenqualität geben:



Diese Punkte müssen bei der Einschätzung der Datenqualität beachtet werden:

- Vollständigkeit: Schätzung der Anzahl, Vollständigkeit einzelner Felder prüfen
- Herkunft, Quelle
- Aktualität
- Redundanz und Konsistenz: Redundanzen entfernen, Konsistenz sicherstellen

Bei niedriger Datenqualität ist entweder eine Auswertung nicht möglich, oder eine sehr aufwendige Auswertung mit hohen Kosten erforderlich.

### 4 Daten aufbereiten

#### 4.1 Daten in die geeignete Form bringen

Unabhängig von der Ursprungsform, ist das Zielformat immer die elektronische Tabelle. Je nach Ursprungsform ist die Überführung mehr oder weniger aufwendig.

Tabellenfelder haben Datentypen. Diese sollen entsprechend sinnvoll definiert werden um die Konsistenz sicherzustellen, die Auswertung zu erleichtern und spezielle Operationen zu unterstützen. Gängige Datentypen sind: Textfelder, Zahlenfelder, Währung, Datum, Uhrzeit etc.

#### 4.2 Daten zur Verfügung stellen

Zur Übertragung einer Tabelle eignen sich oft Standardformate besser als proprietäre Formate um die Kompatibilität zwischen Programmen sicherzustellen. Gängige Standardformate sind Textdateien, darunter ist das CSV Format das am meisten verwendete Format.

## 5 Daten auswerten

### 5.1 Filter bestimmen

Mit Filter kann man die Datenbasis so eingrenzen, dass nur die für den aktuellen Zweck relevanten Daten angezeigt werden.

In Excel unterscheiden wir zwei Varianten: Autofilter und Spezialfilter.

#### 5.1.1 Excel Spezialfilter

Mit dem Excel Filter kann man Daten nach UND-Kriterien (Zeile) und ODER-Kriterien (Reihe) filtern. Die oberste Reihe der Kriterien-Tabelle muss der gleichen Benennung folgen, wie die der Daten-Tabelle. Jede Zeile enthält dann UND-Kriterien. ODER-Kriterien werden mit den Reihen dargestellt:

**Daten-Tabelle (Vorher):**

Monat	Tag	X	Y
Januar	Montag	2134	34
Februar	Dienstag	1234	6
März	Mittwoch	234	345
April	Donnerstag	9023	2134
Mai	Freitag	3904	234
Juni	Samstag	9854	324
Juli	Sonntag	345	45
August	Montag	29830	645
September	Dienstag	34	78
Oktober	Mittwoch	6435	345
November	Donnerstag	234	878
Dezember	Freitag	134	24

**Kriterien-Tabelle:**

Monat	Tag	X	Y
	Montag	<2500	
			<200

**Daten-Tabelle (Nachher):**

Monat	Tag	X	Y
Januar	Montag	2134	34
Februar	Dienstag	1234	6
Juli	Sonntag	345	45
September	Dienstag	34	78
Dezember	Freitag	134	24

### 5.2 Abfragen mit Berechnung

Auswertungswunsch lassen sich oft nicht direkt aus den Daten herauslesen, obwohl die Informationen indirekt vorhanden wären. In solchen Fällen sind weitere Hilfsmittel von Nöten:

- Sortieren, Gruppieren
- Hilfsspalten-/zellen, Berechnungen, Formeln

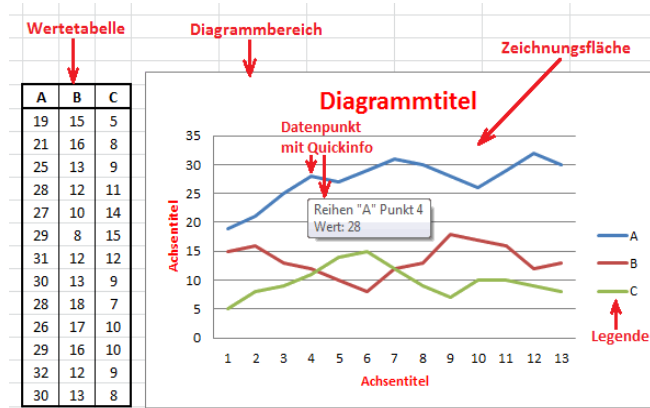


## 6 Daten darstellen

### 6.1 Diagramme erzeugen und gestalten

Diagramme sind eine Abkürzung zur gesuchten Antwort und das beste Mittel, um gezielt auf Sachverhalte aufmerksam zu machen. Sie bieten die Möglichkeit, Fakten abzuschwächen oder zu überzeichnen.

Diagramme lassen sich weitgehend automatisch generieren, jedoch brauchen sie meist noch etwas Nachbearbeitung.



Die Auswahl an Diagrammtypen ist vielfältig.

Nicht alles, was technisch möglich ist, ist auch gewinnbringend.

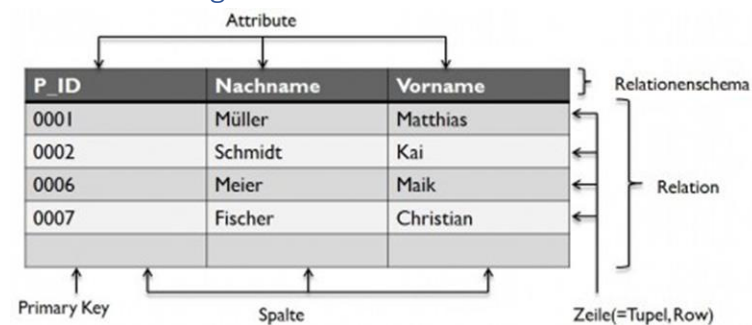
Die aussagekräftigsten Daten können durch eine ungünstige Wahl des Diagrammtyps zunichtegemacht werden.

#### 6.1.1 Ergebnisse durch Darstellung gewichten

Die Darstellung von Sachverhalten kommt grundsätzlich zur Geltung, wenn die betroffenen Werte deutlich auseinanderliegen. Die Gewichtung wird durch den Anzeigebereich, Kurvensteilheit, Auswertgröße und Wertebereich erreicht.

## 7 Relationale Datenbanken

### 7.1 Grundbegriffe



### 7.2 Probleme von normalen Tabellen

- Daten hinzufügen: Redundanzen, Unvollständige Datensätze, Verletzung/Gefährdung der Konsistenz
- Daten ändern: Aufwand, Gefährdung der Konsistenz
- Daten löschen: eventueller Informationsverlust

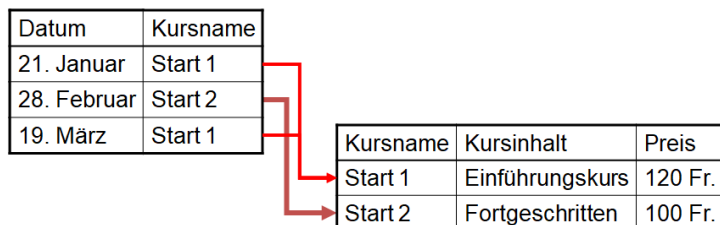


### 7.3 Relationale Datenbanken

Mit Relationalen Datenbanken werden die Probleme von normalen Tabellen (S. [Lernziel 7.2](#)) gelöst:

#### 7.3.1 Redundanzen entfernen

Datum	Kursname	Kursinhalt	Preis
21. Januar	Start 1	Einführungskurs	120 Fr.
28. Februar	Start 2	Fortgeschritten	100 Fr.
19. März	Start 1	Einführungskurs	120 Fr.



#### 7.3.2 Tupel eindeutig identifizierbar machen

Tupel lassen sich durch Primärschlüssel eindeutig identifizieren. Primärschlüssel sind pro Tabelle einmalig, können künstlich generiert sein oder können aus einer Kombination aus Attributen zusammengesetzt werden.

#### 7.3.3 Beziehungen (Assoziationen) zwischen den Relationen herstellen

##### 7.3.3.1 Fremdschlüssel

Fremdschlüssel dienen der Verknüpfung von Tabellen, referenzieren immer einen Primärschlüssel einer anderen Tabelle und gewährleisten, dass Informationen nur einmal abgelegt, jedoch mehrfach verwendet werden können.

##### 7.3.3.2 Beziehungen (Kardinalitäten)

- 1:1 - Jedem Objekt der Art A wird höchstens ein Objekt der Art B zugeordnet.
- 1:n - Jedem Objekt der Art A können mehrere Objekte der Art B zugeordnet werden.
- m:n - Jedem Objekt der Art A/B können mehrere Objekt der Art B/A zugeordnet werden.

