



# Ri

## 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

**Es wurden keine Einträge für das Inhaltsverzeichnis gefunden.**

**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 erkenntungsfähige 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

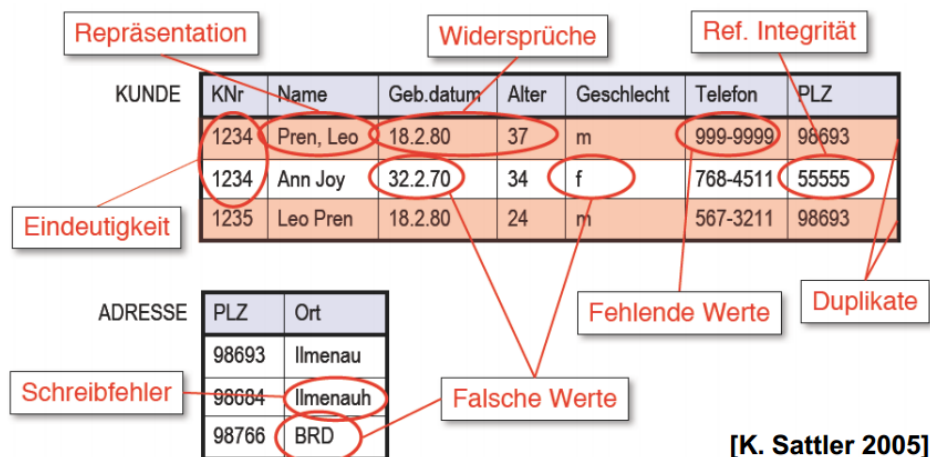
#### 3.1 Daten auswerten

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

