

# Datenbankbasierte Web-Anwendungen

## Einführung in Datenbanken

### Medieninformatik SoSe 2017

Renzo Kottmann



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

# Kontakt:

- mail
- linkedin:<http://www.linkedin.com/in/renzokottmann>
- twitter: @renzokott

# Organisatorisches

Wann machen wir mUes?

auf geht's :)

# Gesellschaftliche Bedeutung von Daten

- Daten werden immer wichtiger
  - Nicht nur bei dot.com Firmen wie:
    - Google, Facebook, Twitter etc.
- Daten sind vom allgemeinem öffentlichen Interesse
  - Begriffe wie 'Big Data', 'Metadata', 'Datenschutz' oder 'Abhörskandal' machen Presseschlagzeilen
- Daten wachsen enorm
  - Schätzung: 1.2 Zettabyte [im Jahr 2010](#)
  - Alle 18 Monate verdoppelt
- Allgemein wachsende Bedeutung in
  - Wirtschaft, Wissenschaft und Politik
  - Stichworte 'Data Economy' und 'Data as a currency'

# Was sind Daten...

- Es gibt mehrere Definitionen zu Daten (engl. data)
  - von denen nur einige frei zugänglich sind:

# Definition: Merriam Webster

data noun plural but singular or plural in construction, often attributive \ 'dā-tə, 'da- also 'dä-\

: facts or information used usually to calculate, analyze, or plan something

: information that is produced or stored by a computer

# Definition: Merriam Webster

## Full Definition of DATA

1: factual information (as measurements or statistics) used as a basis for reasoning, discussion, or calculation

< the data is plentiful and easily available—H. A. Gleason >

< comprehensive data on economic growth have been published — N. H. Jacoby >

2: information output by a sensing device or organ that includes both useful and irrelevant or redundant information and must be processed to be meaningful

3: information in numerical form that can be digitally transmitted or processed

# Definition in der Wirtschaftsinformatik:

Gabler Wirtschaftlexikon Daten im Kontext der Wirtschaftsinformatik:

Zum Zweck der Verarbeitung zusammengefasste Zeichen, die aufgrund bekannter oder unterstellter Abmachungen Informationen (d.h. Angaben über Sachverhalte und Vorgänge) darstellen.

# Definition in der Wirtschaftsinformatik:

Gabler Wirtschaftlexikon Daten im Kontext der Wirtschaftsinformatik:

Zum Zweck der Verarbeitung zusammengefasste Zeichen, die aufgrund bekannter oder unterstellter Abmachungen Informationen (d.h. Angaben über Sachverhalte und Vorgänge) darstellen.

## Daraus kann man folgern, dass Daten

1. semantisch
  - Informationen sind, welche Fakten der realen Welt wiedergeben
2. syntaktisch
  - eine kodierte digitale Folge von Zeichen sind, die nur in einem bestimmten Bedeutungskontext zu einer richtigen Interpretation führen

138

# 138

- Die Zeichen bzw. Zahlenfolge "138" kann je nach Kontext verschiedenes darstellen:
  - eine real existierende Hausnummer oder
  - eine Rechnungsnummer sein oder
  - ein geografischer Breitengrad (z.B. 138 Grad West)
  - oder ...

# Kategorisierung von Daten

Man unterscheidet Datensammlungen häufig anhand des vorliegenden Strukturierungsgrades in:

- unstrukturierte Daten
  - Beispiele: Dokumente, beliebige Texte, Grafiken
- semistrukturierte Daten
  - führen einen Teil der Strukturinformationen mit sich
  - müssen keiner allgemeinen formalisierten Struktur entsprechen
  - Beispiele: Daten gespeichert in Extensible Markup Language (XML), JSON oder CSV
- strukturierte Daten
  - Müssen gemäß einem Datenmodell gleiche Struktur haben
  - Beispiel: relationale Datenbank

**... und was ist nun eine Datenbank?**

# Defintion Datenbank

ist eine zweckorientierte Sammlung von Daten. Daten werden dabei organisiert und strukturiert, um zweckmäßige Aspekte der Welt zu modellieren, sodass die Erfassung und Verarbeitung dieser Daten effizient unterstützt wird.

# Analog und digital

- Analoge Sammlungen von Daten sind Datenbanken!
  - z.B. Aktenschränke oder Karteikartensammlungen
  - Historisch und konzeptionell Vorläufer von modernen elektronischen Datenbanken
- Alle Datenbanken dienen der effizienten
  - Erfassung,
  - Speicherung
  - und Verarbeitung von Daten.

# Nutzen von Datenbanken

- Häufigste Anwendung ist Beantwortung von Anfragen wie z.B.
  - welche Bücher von 'Douglas Adams' sind in dieser Bibliothek erhältlich
  - in welchem Regal befindet sich "Per Anhalter durch die Galaxis"?

# Datenbank Entwicklung

Das Erstellen einer neuen Datensammlung erfordert dazu u.a. die Erfassung von Daten.

# Von unstrukturierten zu strukturierten Daten

- Ein neue Datensammlung beginnt häufig mit der Erfassung der Daten
  - durch Digitalisierung
  - oder liegen schon vor
    - z.B. in einfachen Textdokumenten wie Word-files, PDFs oder auch Internetseiten.

# Erfassung

Daten der Modulteilnehmer

**Digital(isierung)**

**Wie?**

# Digital(isierung)

## Wie?

Zur ersten Erfassung und Verwaltung bietet sich auch ein Tabellenkalkulationsprogramm wie z.B. Excel, OpenOffice oder online Tools wie Google Spreadsheets an.

# Google Spreadsheet

[Google Spreadsheet](#)

URL:

[https://docs.google.com/spreadsheets/d/1tpMyVaZla6N9n7heww\\_y1F78a9S0BRw4NwkI](https://docs.google.com/spreadsheets/d/1tpMyVaZla6N9n7heww_y1F78a9S0BRw4NwkI)

# Erkenntnisse??

# Meine Erkenntnisse:

- Als Datenbankstruktur ist eine Tabelle geeignet:
  - Ein Datensatz pro Zeile
  - Eine Eigenschaft (Attribut) pro Spalte
  - Erste Zeile enthält die Namen der Eigenschaften (anstatt eines Datensatzes)
  - Die Reihenfolge der Zeilen ist egal
  - Die Reihenfolge der Spalten ist egal

# Daten Speicherung

- Files:
  - Nun befinden sich die Daten in einer Datei persistent im Dateisystem gespeichert, d.h. diese werden über die Laufzeit eines Programms oder des Computers hinaus existieren.
  - Nicht zwingend Excel-Format:
  - Excel ist ein binäres Format
  - In vielen Fällen reicht eine Textdatei
    - Z.B. das CSV-Format (Comma-separated values)

# Comma Separated Values

- CSV-Dateien entsprechen Tabellen, gekennzeichnet durch:
  - jede Zeile durch ein Zeilenendezeichen
    - Spalten durch ein Trennsymbol wie
      - z.B. ein Komma ',' oder Semikolon ';'

Eine CSV-Datei kann man mit allen Textverarbeitungsprogrammen und auch Tabellenkalkulationsprogrammen bearbeiten werden.

## Probleme dieser einfachen Struktur:

```
Viereck;Axel;26123;Oldenburg;  
Huber;Ina;12345;FFM;0123/65235  
Lustig;Olga;12345;Frankfurt;0123/45456  
Mustermann;Erika;12345;Frankfurt;0123/45456  
Henseler;Herwig;26197;Großenkneten;04435/388486 (Fax:388487)  
Lustig;Peter;Frankfurt;0123/45456  
Huber;Ina;3454;Dresden;0283/11111  
Mustermann;Erika;;Bremen;436654
```

## Probleme dieser einfachen Struktur:

```
Viereck;Axel;26123;Oldenburg;  
Huber;Ina;12345;FFM;0123/65235  
Lustig;Olga;12345;Frankfurt;0123/45456  
Mustermann;Erika;12345;Frankfurt;0123/45456  
Henseler;Herwig;26197;Großenkneten;04435/388486 (Fax:388487)  
Lustig;Peter;Frankfurt;0123/45456  
Huber;Ina;3454;Dresden;0283/11111  
Mustermann;Erika;;Bremen;436654
```

- Mögliche Redundanzen
- Beziehungen werden nicht repräsentiert
- Keine Festlegung von Datentypen und Datenintegritätsbedingungen
- Unklare Eindeutigkeiten

# Semistrukturierte Daten: XML

```
<?xmlversion="1.0"?>
<adressen>
  <adresse>
    <nachname>Lustig</nachname>
    <vorname>Peter</vorname>
    <plz>12345</plz>
    <ort>Frankfurt</ort>
    <telefon>0123/45456</telefon>
  </adresse>
  <!-- einige Eintraege ausgelassen -->
  <adresse>
    <telefon>436654</telefon>
    <nachname>Mustermann</nachname>
    <vorname>Erika</vorname>
    <ort>Bremen</ort>
  </adresse>
</adressen>
```

# Semistrukturierte Daten: JSON

```
{
  "adressen": [{
    "nachname": "Lustig",
    "vorname": "Peter",
    "plz": "12345",
    "ort": "Frankfurt",
    "telefon": "0123/45456"
  },
  {
    "telefon": "436654",
    "nachname": "Mustermann",
    "vorname": "Erika",
    "ort": "Bremen"
  }
  ]
}
```

# Semistrukturierte Daten

- Für einfache Anwendungen kann diese Form der Datenspeicherung und verwaltung in Datei(en) durchaus ausreichen.
- Filesysteme unterstützen nicht (oder nur unzureichend):
  - effiziente Suche und Modifikation von kleinen Dateneinheiten,
  - komplexe Datenanfragen,
  - Transaktionen
  - effizientes buffering und caching von Daten im Hauptspeicher

Anforderungen an Datenbanken in der Regel deutlich höher.

# Schritte im ad-hoc Projekt

Datenverwaltungssystem: semistrukturierte Daten im Filesystem

## 1. Problem in der realen Welt

- Wer sind die Teilnehmerinnen in meinem Kurs?

## 2. Datenerfassung

- Liste der Teilnehmerinnen
- Implizites Model (Tabelle)

## 3. System zur Verarbeitung und Speicherung

- CSV Datei

**Was sind die Nachteile von diesem System  
und dieser Vorgehensweise?**

# Was wäre besser?

# Was wäre besser?

1. Hohe Datenintegrität bzw. Konsistenz
2. Mehrere Nutzer können gleichzeitig an den selben Daten arbeiten
3. Klare Trennung von Struktur und Inhalt bzw. Daten
4. Effiziente Suche in grossen Daten

# Für Datenintegrität und Strukturtrennung

- Strukturierte Daten
- (relationale) Datenbank (DB)
  - Schema („Modell“, „Metadaten“)
  - + Daten („Extension“, „Nutzdaten“)
  - Modell und Daten sind konzeptionell getrennt!

# Für Effizienz und Mehrbenutzer

- Datenbankmanagementsystem (DBMS)
- Softwaresystem
  - Ermöglicht Erstellung und Pflege vieler Datenbanken
  - Stellt Werkzeuge für alle Aspekte der Datenverwaltung bereit

## 2. Runde

Mit dem relationalem Ziel vor Auge

# Schritte im ad-hoc Projekt

Datenverwaltungssystem: semistrukturierte Daten im Filesystem

## 1. Problem in der realen Welt

- Wer sind die Teilnehmerinnen in meinem Kurs?

## 2. Datenerfassung

- Liste der Teilnehmerinnen
- Implizites Model (Tabelle)

## 3. System zur Verarbeitung und Speicherung

- CSV Datei

# Schritte im agilen Datenbank Projekt

Datenverwaltungssystem: Relationales Datenbankmanagementsystem

## 1. Anwendungsdefinition

- Genaue textuelle Beschreibung des Problems in der realen Welt.
- Wer sind die Teilnehmerinnen in meinem Kurs?

## 2. Konzeptioneller Entwurf / Datenbankmodellierung

- Entity Relationship Modellierung (ERM)

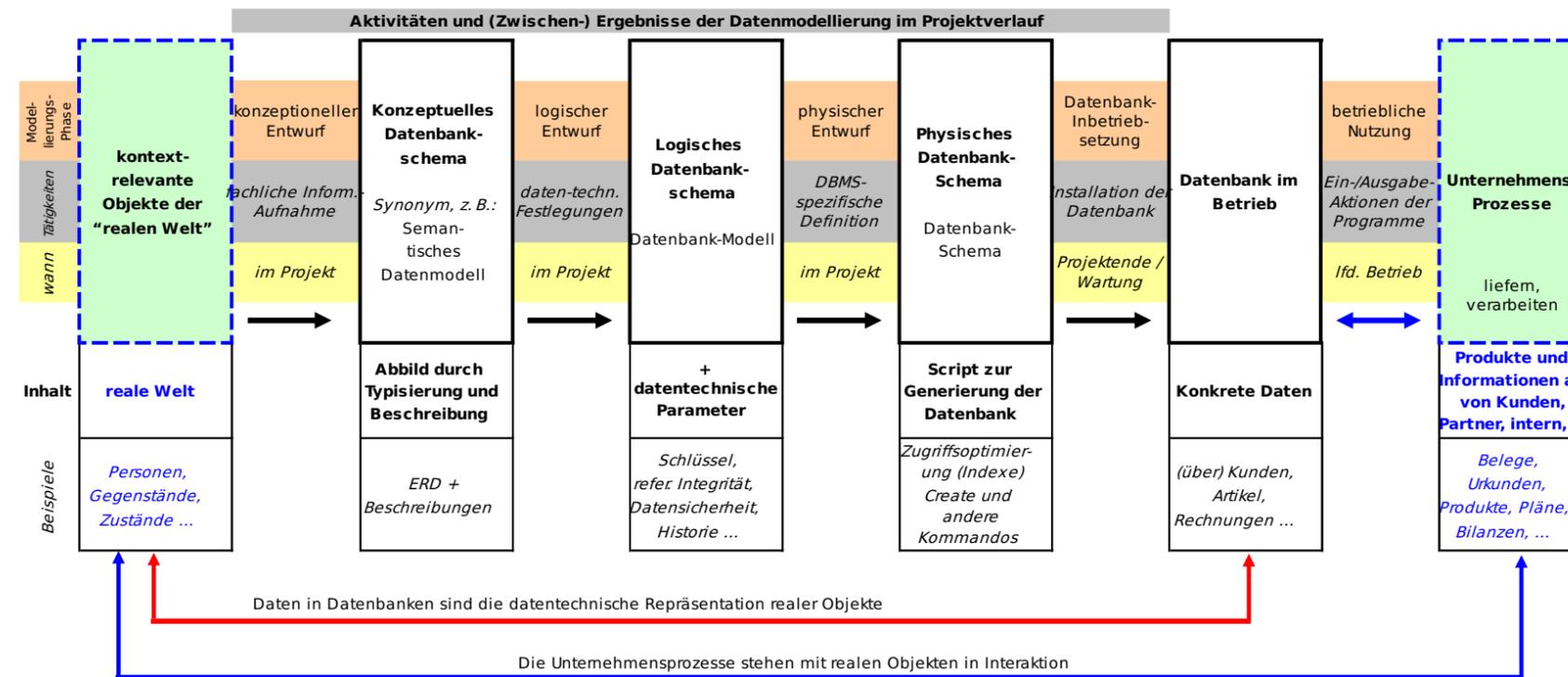
## 3. Datenbank Realisierung a) Skript(e) zur Implementierung des Datenmodells b) Skript(e) zum laden/einfügen der Daten

## 4. Von Geschäftsanfragen zu Datenbank-Abfragen

- Skript(e) für SQL ´s SELECTs

# Alternativer Blick: Datenbankentwicklungszyklus

Modellieren: Entwicklung von der fachlichen, implementierungsunabhängigen Konzeption bis zur Datenbank



„[DatMod v semMod zur DBK](#)“ von [VÖRBY](#), Konvertierung zu SVG [Perhelion](#) - eigene Erstellung, aus Wikipedia-Text abgeleitet. Lizenziert unter Gemeinfrei über [Wikimedia Commons](#).

# Anwendungsdefinition

In der Phase der Systemanalyse werden ausgehend von der Problemstellung die Anforderungen an die Lösung formuliert. Diese sollten möglichst vollständig und konsistent sein, d.h. alle (vollständigen) Anforderungen sollten widerspruchsfrei (konsistent) formuliert werden.

Ein guter Einstieg in die Anwendungsdefinition ist, sich zu verdeutlichen, welche genaueren Zwecksetzungen erfüllt werden sollen. D.h. was ist das Ziel des Projektes.

# Anwendungsdefinition

Welchen Zwecken soll die Erfassung aller Kursteilnehmerinnen dienen?

# Anwendungsdefinition

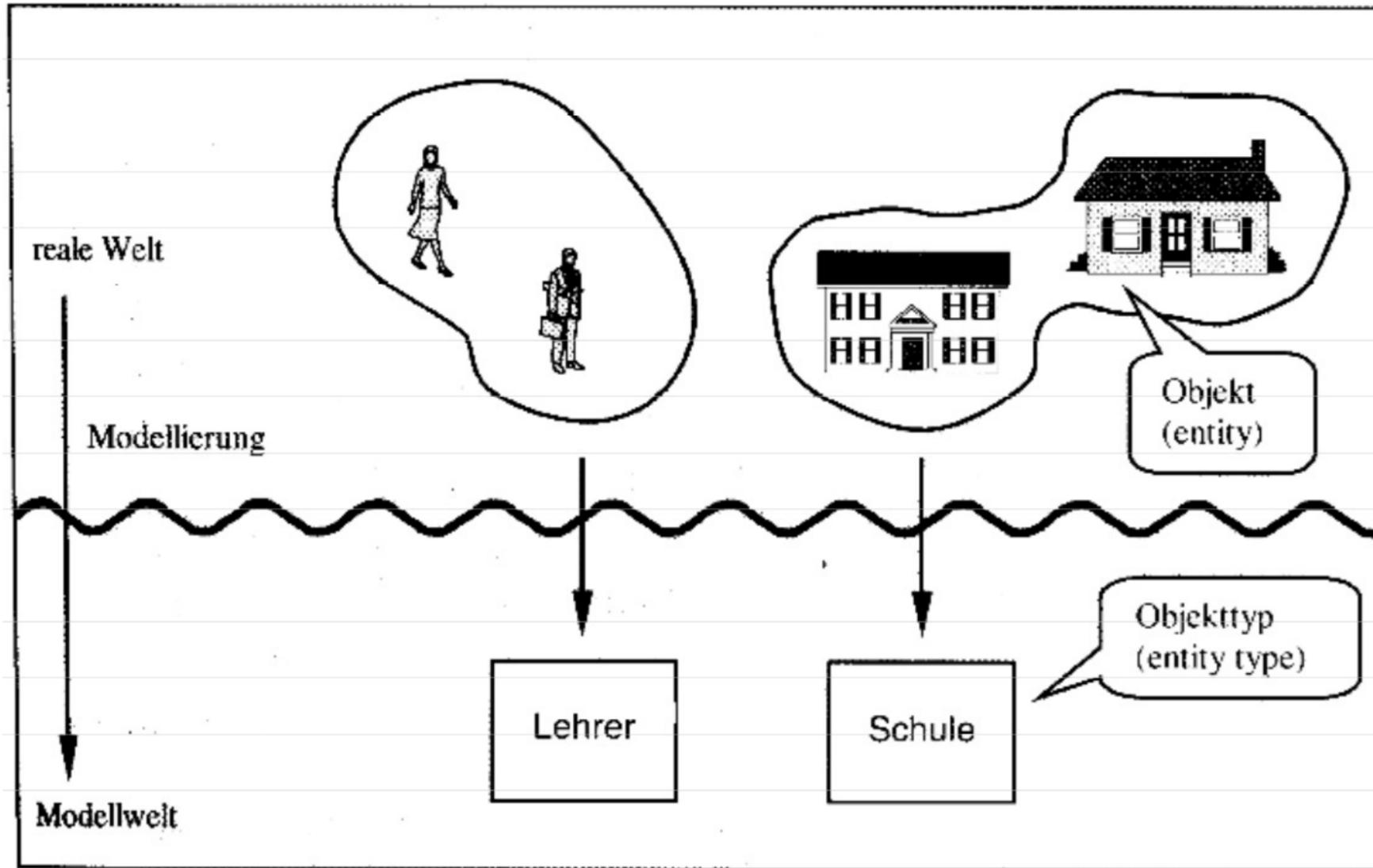
Welchen Zwecken soll die Erfassung aller Kursteilnehmerinnen dienen?

- Informationsweitergabe an alle Teilnehmerinnen
- Namen lernen
- Notenvergabe
- Ermittlung des Wissensstands
- Wer macht welches Datenbankprojekt
- Wer ist in welcher Laborgruppe

# Mini-Welt

(wenn) wir wissen was wir wollen, dann können wir folgende Fragen beantworten:

- Welchen Ausschnitt der realen Welt brauchen wir?
- Welche Aspekte müssen wir Berücksichtigen?



# Entity Relationship Modellierung (ERM)

- Semantischer Datenbankentwurf
  - unabhängig von konkreten Datenbank-spezifischen Modellen
- Graphisch
- Idee simpel
  - Leider sehr viele inkonsistente Varianten

# Entity (Entität)

- Ein Entity-Relationship-Model (ERM) geht von Entitäten ( $\approx$  Objekten) aus.

"Eine Entität ist eine eigenständige Einheit, die im Rahmen des betrachteten Modells **eindeutig identifiziert** werden kann."

- Ein Entitätstyp wird durch Attribute genauer beschrieben und stellt somit eine abstrakte Beschreibung oder Charakterisierung von Entitäten da.

- Beispiel:

Lehrer = Entitätstyp

Renzo  $\approx$  Entität (ein spezieller Lehrender)

# Attribute

- Eigenschaften von Entitäten werden durch Attribute beschrieben
- Attribute haben einen Namen und eine Domain (= Bestimmung der Wertmenge).

# Keys (Schlüssel)

Da die Definition einer Entität beinhaltet, dass diese zumindest im Rahmen eines Modells eindeutig identifiziert werden kann, braucht jeder Entitätstyp eine Menge von Attributen als Schlüssel.

Die Auswahl eines oder mehrerer Attribute als Schlüssel legt fest, dass es keine zwei Entitäten eines Entitätstyp geben kann die identische Attributwerte haben.

- Wichtige Eigenschaften:
  - Eindeutigkeit
  - Zuteilbarkeit

# Relationship (Beziehung)

Verschiedene Entitäten können zueinander in Beziehung gesetzt werden.

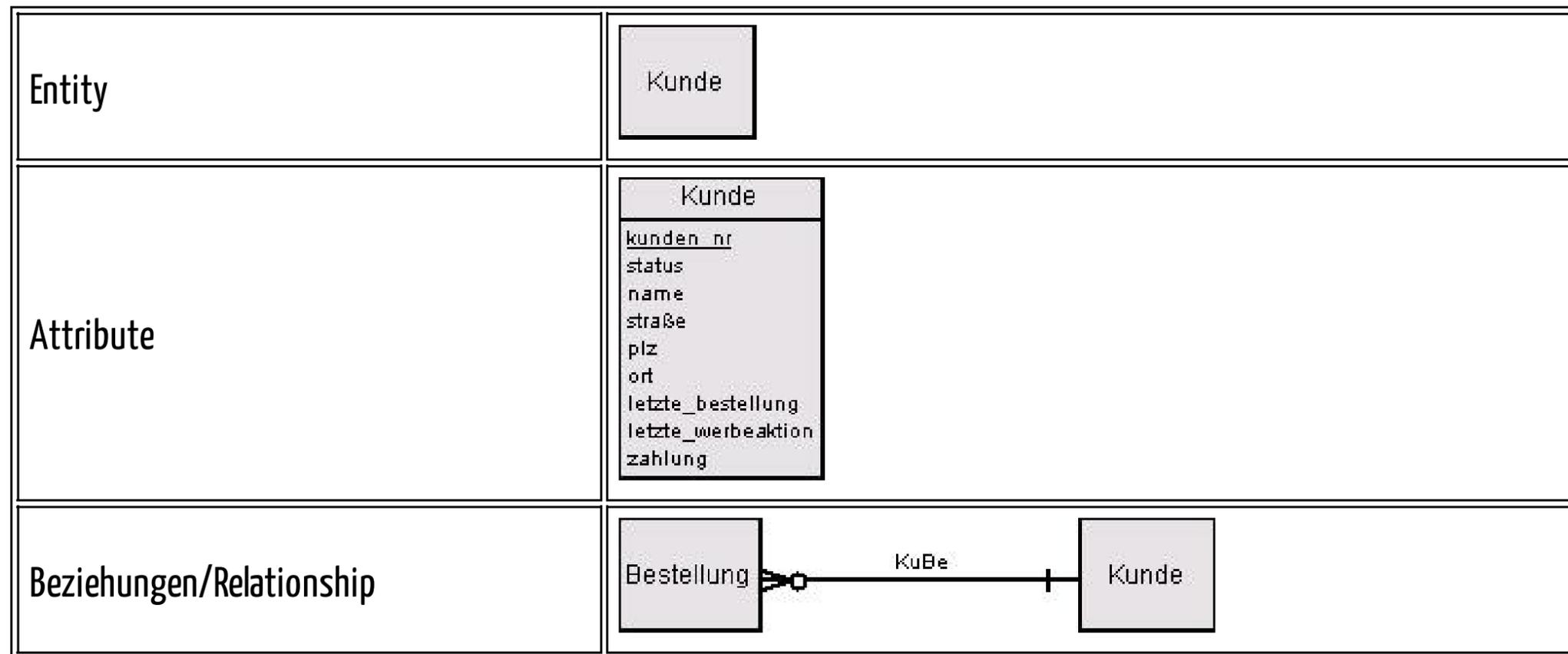
- In jeder Beziehung haben Entitäten gewisse Rollen
- Beziehungen können Eigenschaften (Attribute) haben
- Beziehungen haben Kardinalitäten

# Notationen

Es gibt verschiedene Formen ERM zu notieren (textuell und/oder graphisch):

- [Chen Notation](#)
- [Crow Foot's](#)
- [Unified Modelling Language \(UML\)](#)

# Notation



**Danke fuer die Zusammenarbeit**