

Datenmodellierung und Analyse

Kursziele

- 1) *Daten-Modelle* kennen und anwenden können
- 2) *Datenbank-Grundlagen* verstehen und auf eine MySQL-Datenbank lesend zugreifen können
- 3) Mit diversen *Daten-Formaten* umgehen können
- 4) Die Mittel kennen, um Daten zu *suchen*, *analysieren* und *visualisieren*

Kursinhalt (Tage)

1. Einleitung / das ER-Modell
2. ER-Modell, Datenbanktypen, SQL Einleitung
3. SQL Survival Kit
4. Datennormalisierung / Transaktionen
5. Die Datenformate XML, JSON und CSV, Zeichenkodierung
6. Repetition, Prüfung, BYOP
7. Datenanalyse mit Jupyter Notebooks (1)
8. Datenanalyse mit Jupyter Notebooks (2)
9. Datenanalyse mit Jupyter Notebooks (3)

Tag 1

Inhaltsverzeichnis

- Literaturverzeichnis
- Was / wozu Datenmodellierung?
- Datenmodelle
- Das Entity-Relationship Modell
- Übungen
- BYOQ (nach "lean coffee"-Ansatz)

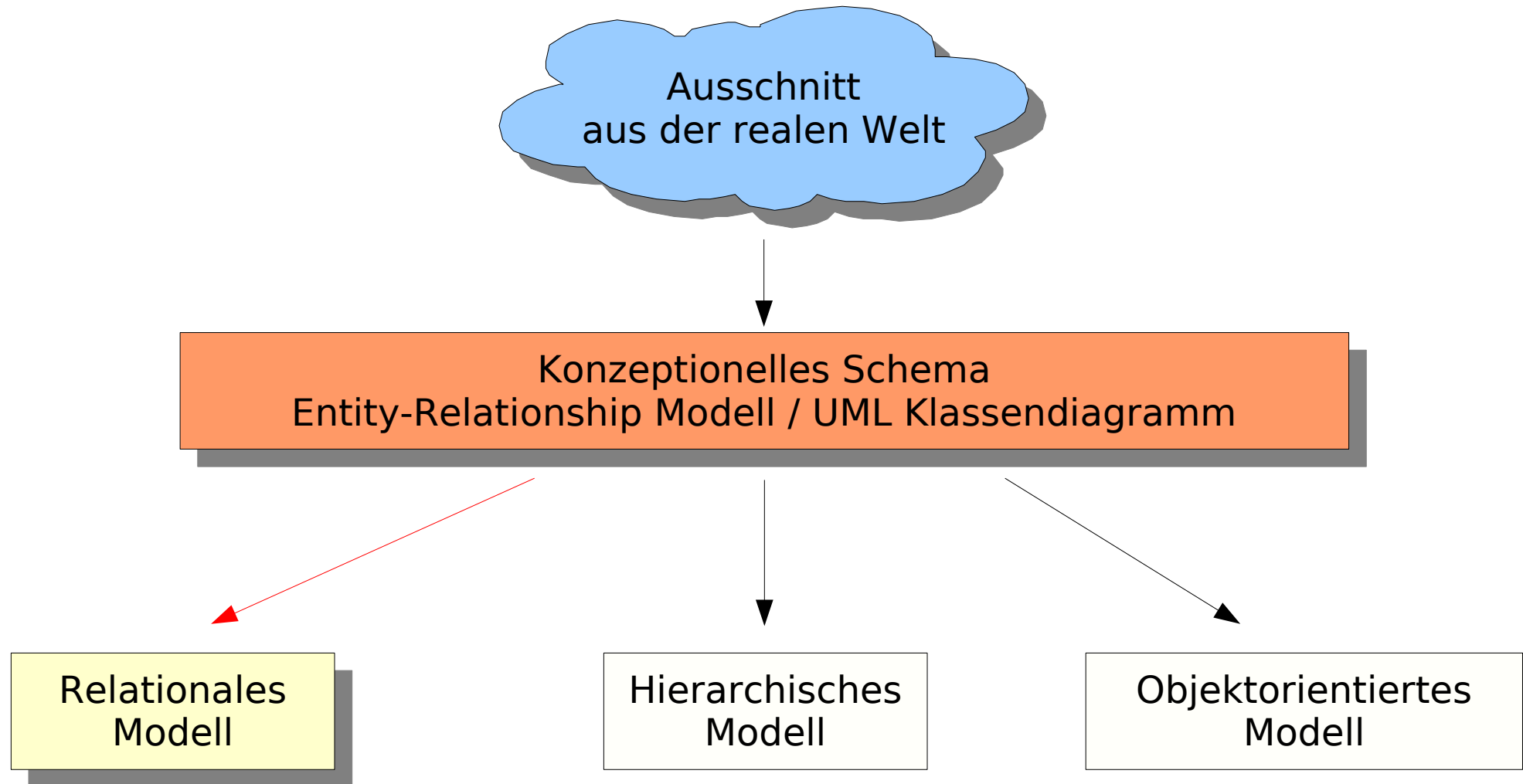
Literaturverzeichnis

- Taschenbuch Datenbanken / Thomas Kudraß / Hanser 2015
- Next Generation Databases / G. Harrison / APRES IOUG 2015
- Datenanalyse mit Python: Auswertung von Daten mit Pandas, NumPy und Ipython / Wes McKinney / O'Reilly 2018
- [O'Reilly online learning platform](#)
 - Die grössten US und deutschen Verlage machen mit
 - > 30'000 IT-Bücher und Videos online (auf deutsch und englisch)
 - Privates Starter-Kit für \$39 / Monat

Was / wozu Datenmodellierung?

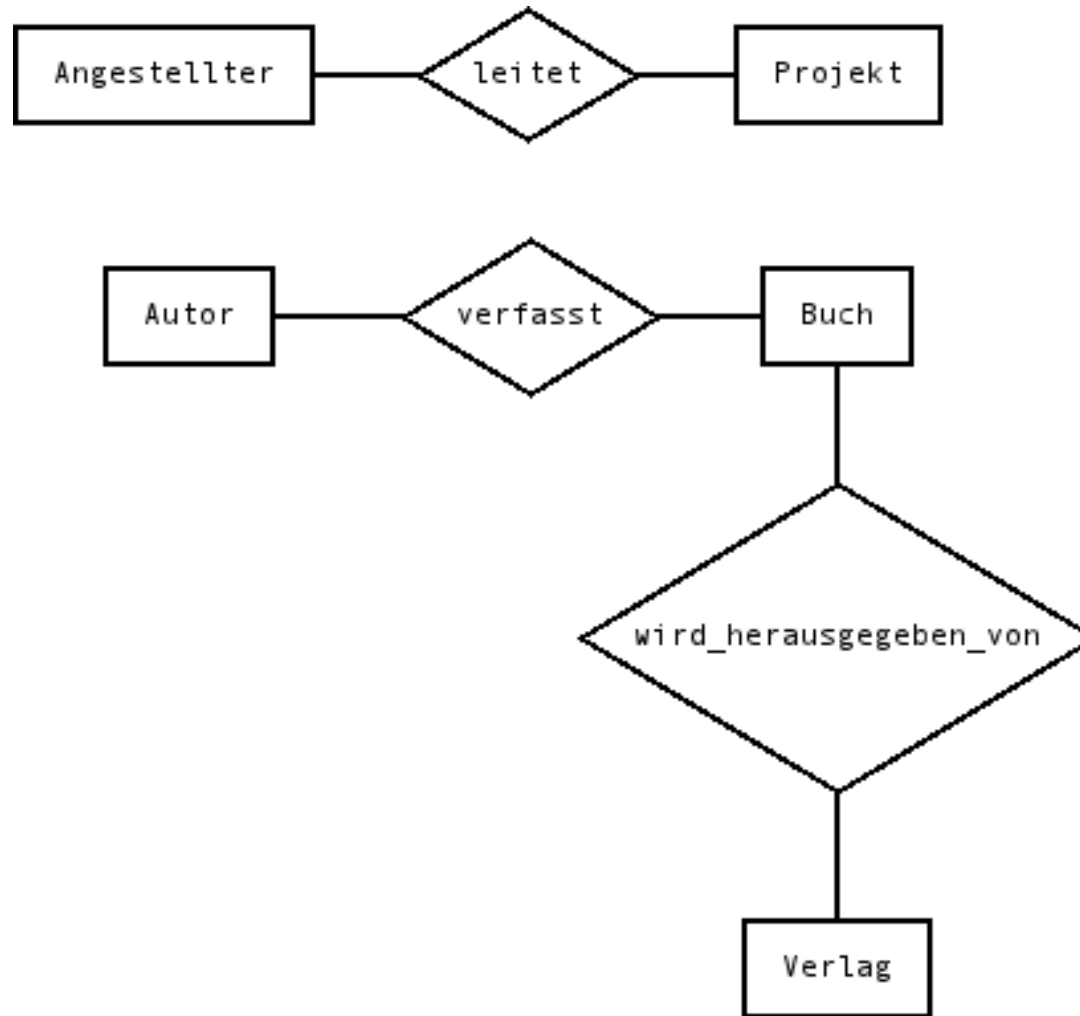
- Daten leben länger als Programme
- "Know your data"
- Daten und ihre Attribute untersuchen
- Daten und ihre Zusammenhänge verstehen
- => Daten abstrahieren => Modell erstellen

Daten- und Datenbankmodelle

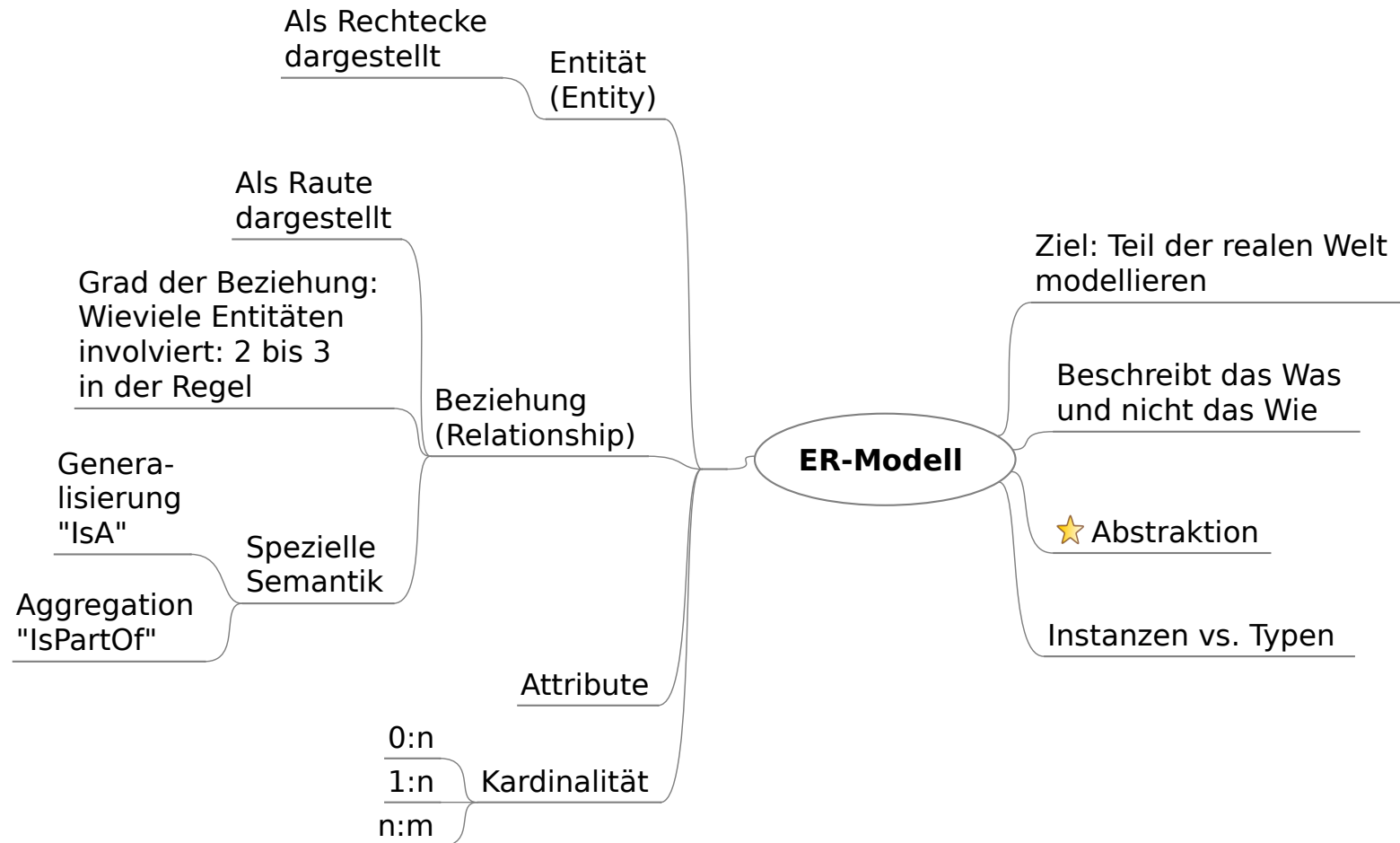


Das relationale Datenmodell

Entity-Relationship Modell / Grundprinzip

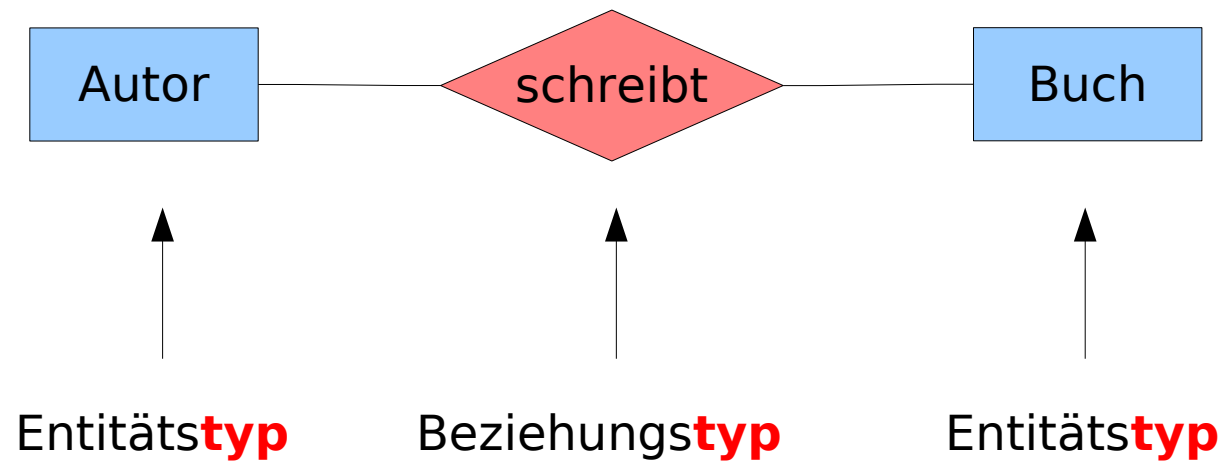


Das Entity-Relationship Modell (ERM) Überblick



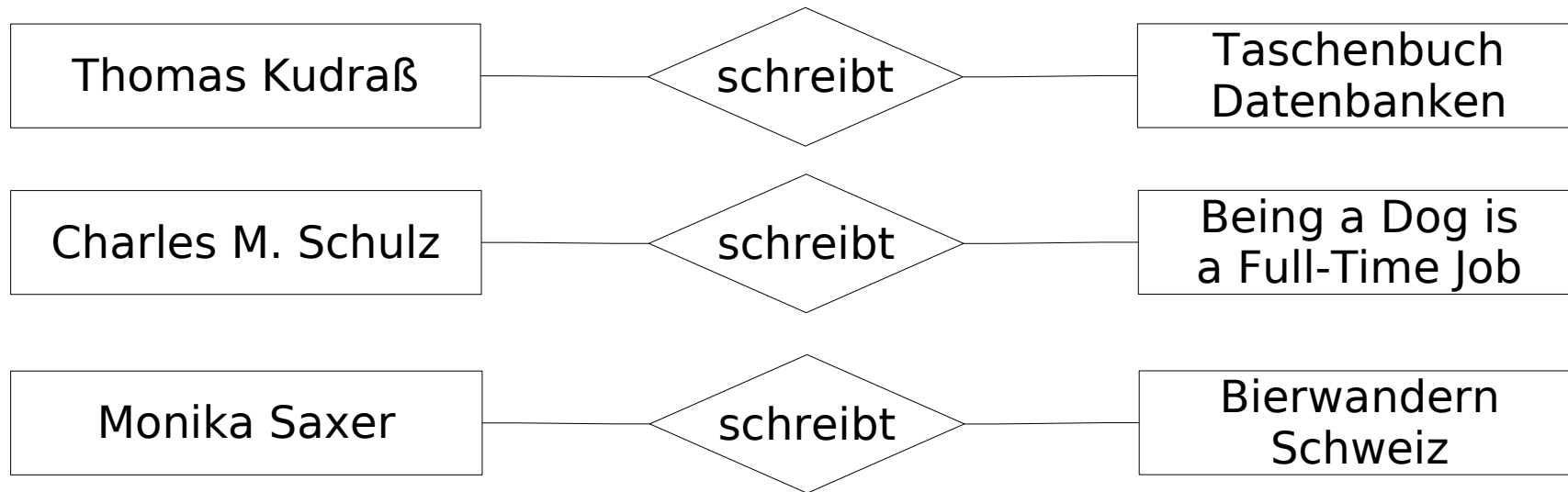
Das ER-Modell

Entitäts- und Beziehungstypen



Das ER-Modell

Entitäts- und Beziehungsinstanzen

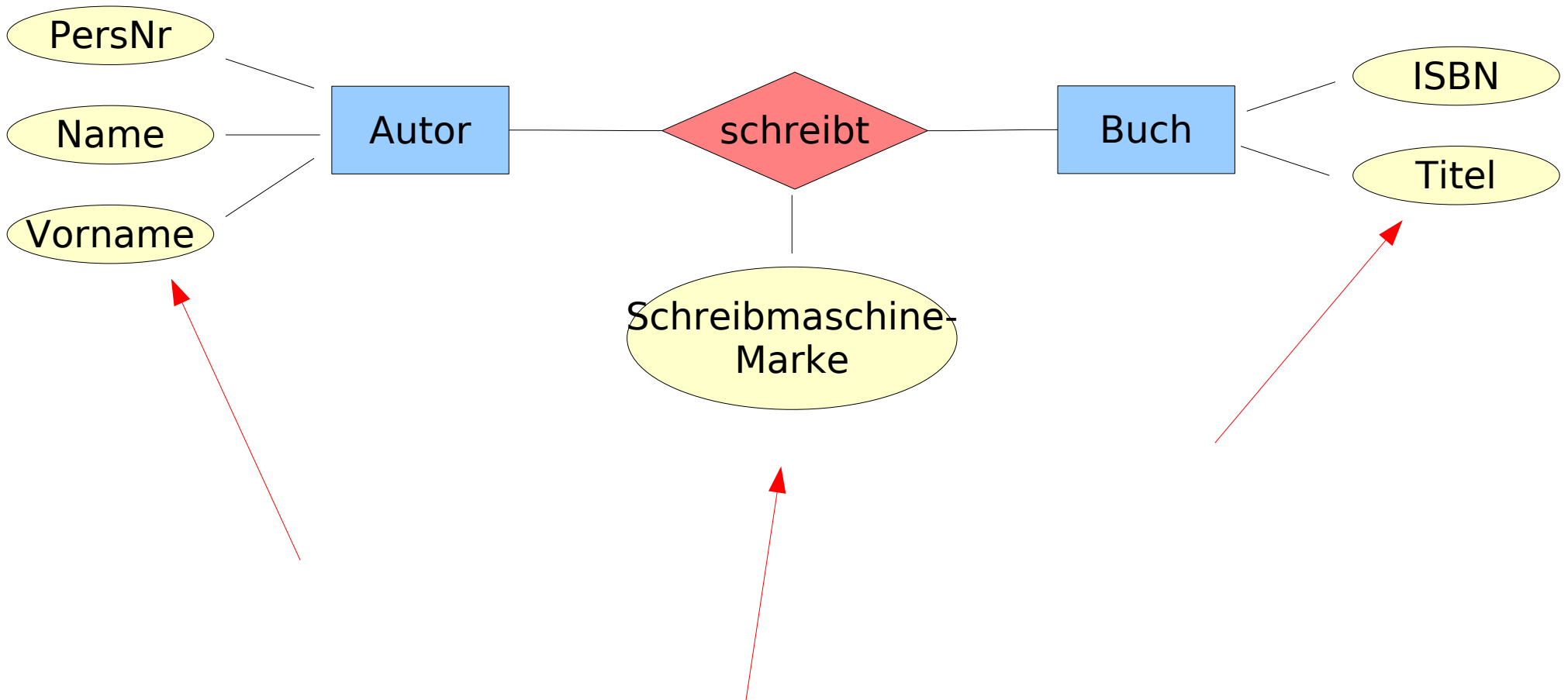


↑
Entitäten

↑
Beziehungen

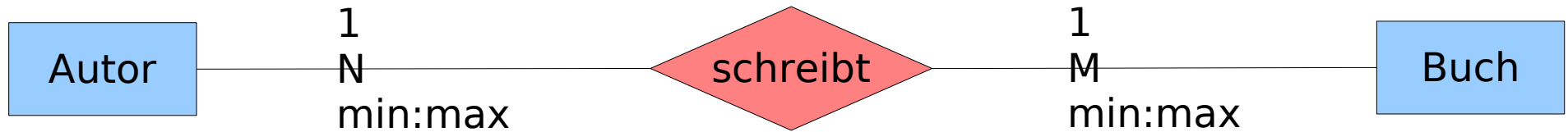
↑
Entitäten

Das ER-Modell Attribute



Das ER-Modell

Kardinalität (1)

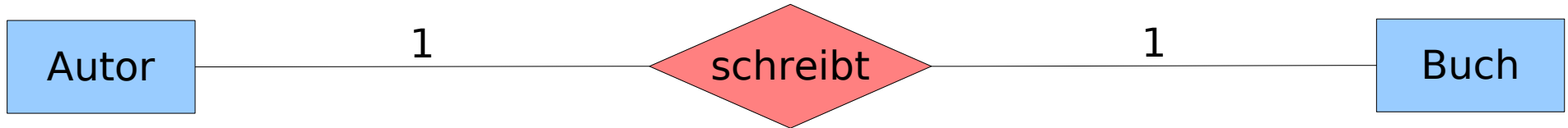


Typische Fragen:

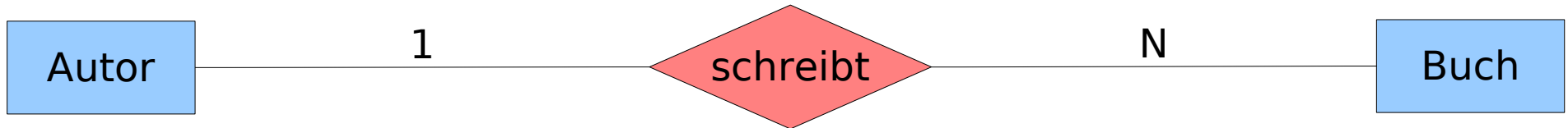
- Gibt es Autoren, die kein Buch geschrieben haben?
- Gibt es Bücher, die von keinem Autor geschrieben werden?
- Gibt es Autoren, die mehr als ein Buch schrieben?
- Gibt es Bücher, die durch mehr als einen Autor geschrieben wurden?

Das ER-Modell

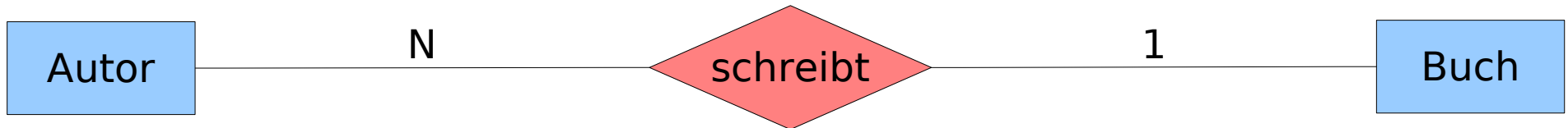
Kardinalität (2)



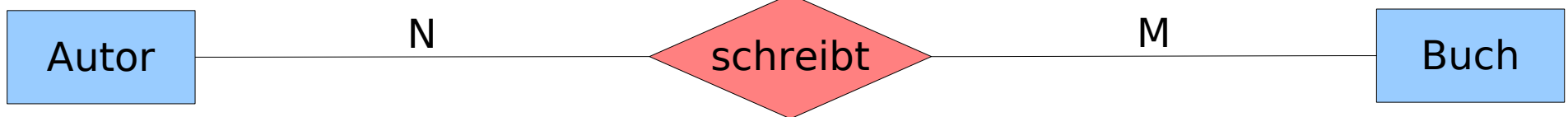
1 Autor schreibt 1 Buch



1 Autor schreibt 0 bis N Bücher



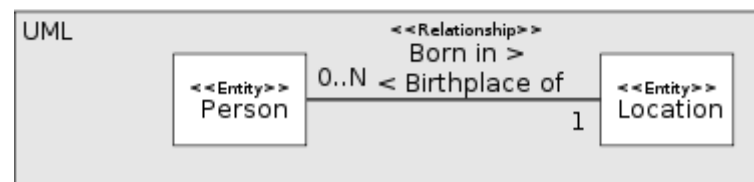
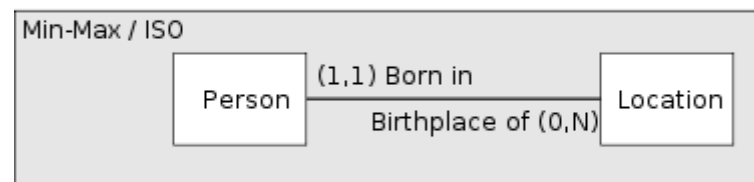
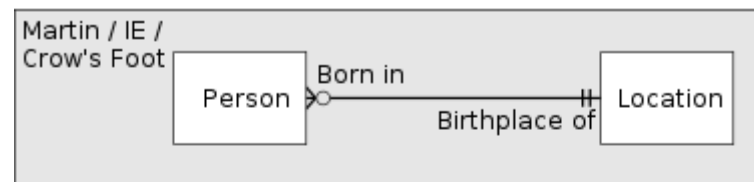
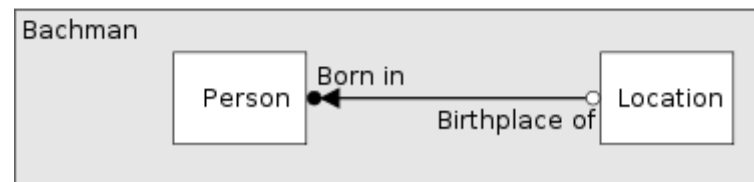
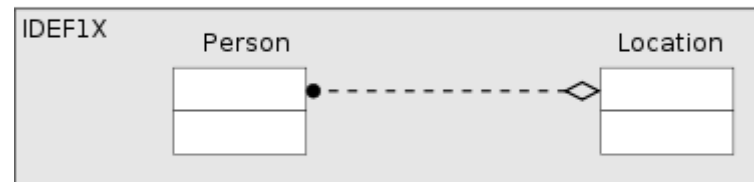
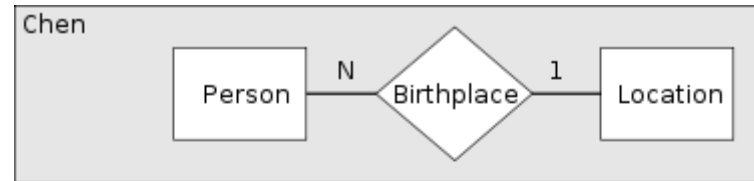
0 bis N Autoren schreiben 1 Buch



0 bis N Autoren schreiben 0 bis M Bücher

Das ER-Modell

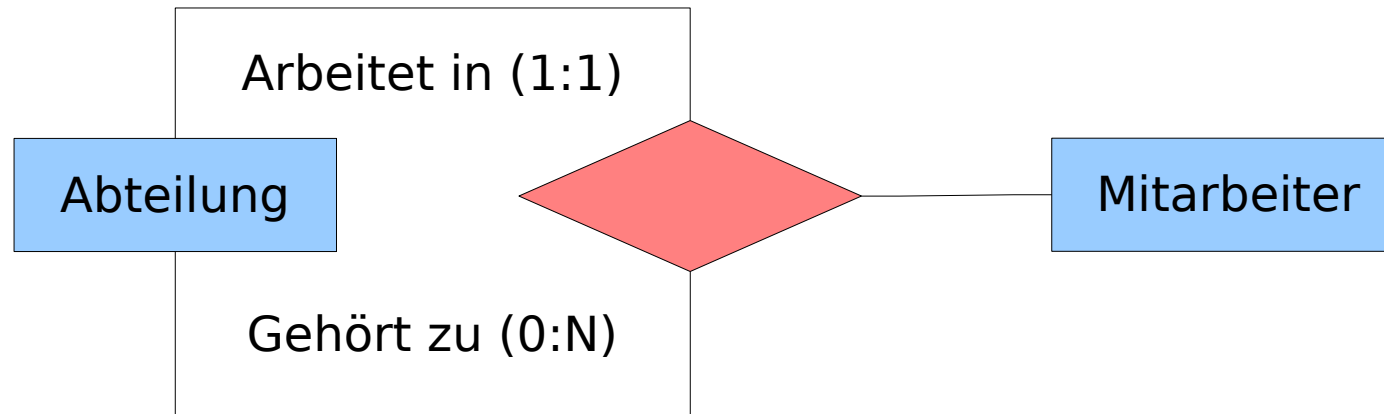
Kardinalität (3) Notation



Quelle: Wikipedia

Das ER-Modell

Umgang mit der Rekursion

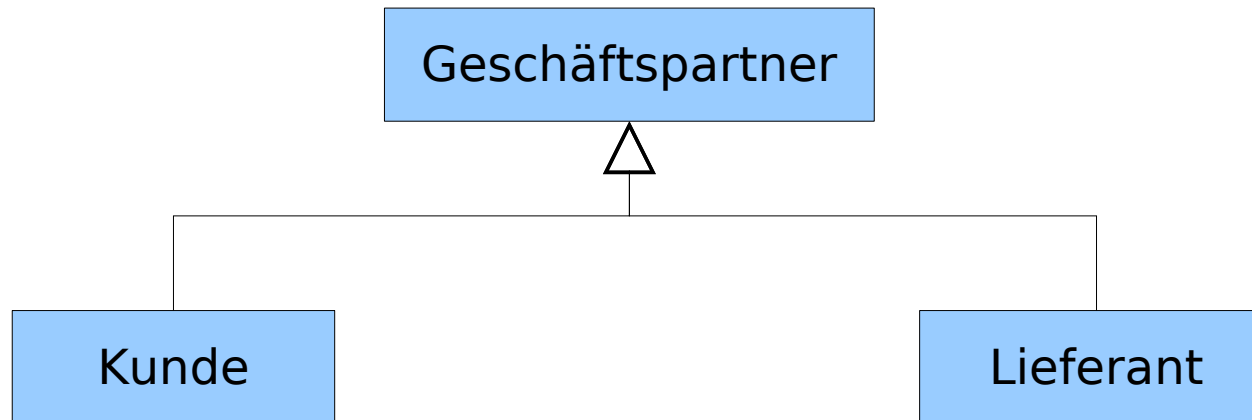


Beispiel

Der Mitarbeiter "*Düschmoll*" arbeitet in der Abteilung "*Software-Engineering*", die selber zur Abteilung "*Entwicklung*" gehört.

Das ER-Modell

Umgang mit der Generalisierung



Typischer Fall eines "Is-A"-Beziehungstyps

=> Zum Beispiel:

Beide *Kunde* und *Lieferant* sind *Geschäftspartner*

Das ER-Modell

Vorgehen für die Diagrammentwicklung

Wichtig

- *Iterativ* vorgehen
- Modell mit Stakeholders *abstimmen*

- (1) Nach *Substantiven* suchen, um die Entitätstypen und deren Attribute zu identifizieren
- (2) Nach *Verben* suchen und damit die Beziehungstypen identifizieren
- (3) Entitäts- und Beziehungstypen *verknüpfen*
- (4) Die *Kardinalität* für jedes Paar "Entitätstyp / Beziehungstyp" spezifizieren

Übung 1

Software-Firma ER-Modellierung

- Die Software Firma "Napadnom" zählt 30 Angestellte. Diese arbeiten in diversen Projekten. Die Projekte brauchen verschiedene Grundsoftware und Hardware, um realisiert zu werden.
- Gewünscht ist ein Informationssystem, das der Firma erlaubt, Projektressourcen zu managen.
- *Tipp 1: Lösen Sie diese Übung in Zweiergruppen*
- *Tipp 2: Auf einem Blatt Papier mit Bleistift und Radiergummi...*

Übung 2

Softwares bereits installiert?

- MySQL Workbench

Client-Software (MySQL spezifisch), um auf DB mit SQL zuzugreifen, DB zu verwalten (kreieren, ändern, löschen) und zu designen (grafisches Schema-Editor)

- Anaconda3

Eine (grosse) Menge Software, um Daten professionell zu analysieren und visualisieren. Die Ziffer 3 deutet, es sei die Version für Python 3.

Siehe [meine Homepage](#) / CAS ITP für mehr Information.

Übung 3

ER-Modell Ihrer CD-Sammlung

- Erstellen Sie das ER-Modell Ihrer CD-Sammlung

Zuerst nur Entitäten, Relationen und Kardinalität modellieren. Dann die Attribute addieren.