

# Tutorium 2

## Relationales Modell, Relationale Algebra

### Big Data Engineering

Prof. Dr. Jens Dittrich

[bigdata.uni-saarland.de](http://bigdata.uni-saarland.de)

9./10. Mai 2022

# Verbesserung Übungsblatt 2 - Häufige Fehler

## Aufgabe 1:

- 1-N- und 1-1-Beziehungen nicht vereinfacht.
- Falsche Vereinfachungen (z.B. den Entitätstypen Personen weggelassen und die Attribute angebracht bei Busfahrer\*innen etc.).
- Falsche Notation bei Fremdschlüsseln

## Aufgabe 2:

- Falsche Annahmen bzgl. des Relationenschemas (z.B. dass nur ein\*e Lehrer\*in alle Schüler\*innen einer Klausur korrigiert).
- Fehlende Umbenennungen.
- Kein Gebrauch von syntaktischem Zucker (z.B. anstatt  $\pi_{ID, Name}$  wurde  $\pi_{\{[ID:int, Name:str]\}}$  verwendet). Dies ist zwar nicht falsch, aber umständlich.
- Keine gleichen Schemata bei Mengenoperationen.
- Statt einem Join Konstrukte wie z.B.  $\sigma_{Attribut = Relation}$  oder  $\in$ .
- 2.6: Join von Lehrer\*innen auf unterrichten. Hierdurch gehen Lehrer\*innen verloren, die zwar Mathe als Hauptfach haben, aber noch nie unterrichtet haben.

# Relationale Algebra - Das heutige Modell

[Personen] : {[PID:int, Name:varchar, Wohnort: varchar, Geburtsjahr: int]}

[Sänger\*innen] : {[SID:(Personen→PID), Künstler\*innenname:varchar, Genre:varchar]}

[Musiklabels] : {[MID:int, Name:varchar, Kapital:varchar, Gründungsjahr: int]}

[Songs] : {[SongID:int, Label:(Musiklabels→MID), Titel: varchar,  
Veröffentlichungsdatum: date]}

[singen\_live] : {[Song:(Songs→SongID), Sänger\*in:(Sänger\*innen→SID), Datum:date,  
Arena:string]}

# Relationale Algebra - RelaX Hands-On

## RelaX - Relationale Algebra Calculator

Implementierung der relationalen Algebra mit Webinterface  
von Johannes Kessler (Uni Innsbruck)

Bereits von uns erstellte Datensätze:

IMDb\_sample

fotodb

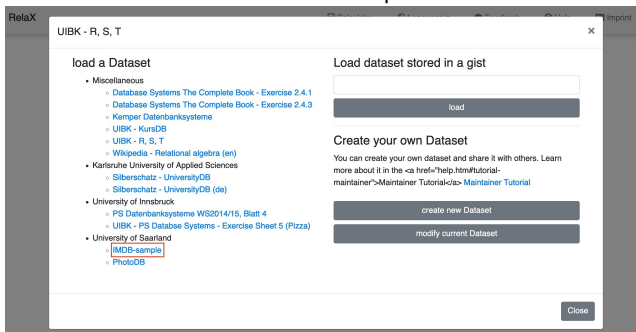
Ihr könnt gerne weitere Datensätze erstellen.

# Relationale Algebra - Relax Hands-On

Geht folgendermaßen vor um einen bestehenden Datensatz mit RelaX zu nutzen:

- öffnet zunächst RelaX und klickt auf das Dropdown-Menü Select DB oben links
- wählt nun unter load a Dataset und University of Saarland den entsprechenden Datensatz aus

Das sollte beim IMDb\_sample so aussehen:



# Relationale Algebra - RelaX Hands-On

Nach erfolgreichem Laden wird das Schema des Datensatzes auf der linken Seite angezeigt und ihr könnt anfangen Ausdrücke der relationalen Algebra darauf auszuwerten.

RelaX - relational algebra calculator 0.19.1

Language ▾ Take a Tour Feedback Help

Success: gist loaded successfully

IMDB-sample ▾

Relational Algebra SQL

Group Editor

actors

id number  
first\_name string  
last\_name string  
gender string

directors

id number  
first\_name string  
last\_name string

directors\_genres

director\_id number  
genre string  
prob number

movies

id number  
name string  
year number  
rank number

movies\_directors

director\_id number  
movie\_id number

movies\_genres

movie\_id number  
genre string

roles

actor\_id number  
movie\_id number  
role string

1 | your query goes here ...

keyboard shortcuts:  
execute statement: [CTRL]+[RETURN]  
execute selection: [CTRL]+[SHIFT]+[RETURN]  
autocomplete: [CTRL]+[SPACE]

▶ execute query

download history ▾

Source: [gist.github.com](https://gist.github.com/jensdiltrich)  
by [jensdiltrich](#) (1 comment)  
last modified 3 days ago

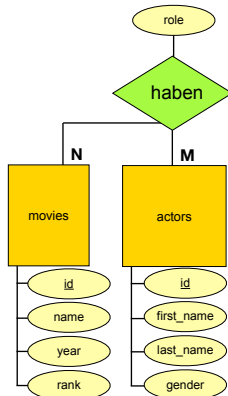
6/32

# Wiederholung - Frage 1

## Frage

Welche Umsetzung der Relation [haben] ergibt im Kontext unseres IMDb ER-Modells mehr Sinn? Begründen Sie Ihre Antwort.

- (a) [haben] : {[ movie\_id:(movies→id), actor\_id:(actors→id), role:str ]}
- (b) [haben] : {[ movie\_id:(movies→id), actor\_id:(actors→id), role:str ]}





Dr. Strangelove oder: Wie ich lernte, die Bombe zu lieben (1964)

Edit

## Full Cast &amp; Crew

See agent

## Directed by

Stanley Kubrick




## Writing Credits

Stanley Kubrick

Terry Southern

Peter George

Peter George

Cast (in credits order) verified as complete Peter Sellers ... Group Capt. Lionel Mandrake / President Merkin Muffley / Dr. Strangelove George C. Scott ... Gen. 'Buck' Turgidson Sterling Hayden ... Brig. Gen. Jack D. Ripper Keenan Wynn ... Col. 'Bat' Guano Slim Pickens ... Maj. 'King' Kong

Dr. Strangelove oder: Wie ich lernte, die Bombe zu lieben

## Details

Full Cast and Crew

## Lösung

## ACHTUNG!

Peter Sellers hatte in "Dr. Strangelove" **drei** verschiedene Rollen!  
Deswegen brauchen wir hier die alternative Modellierung:

[haben] : { [ movie\_id:(movies→id), actor\_id:(actors→id), role:str ] }

Da dies natürlich prinzipiell in jedem Film vorkommen kann, muss die Relation [haben] so modelliert werden.

## Related lists from IMDb users



## 1960s

a list of 29 titles  
created 29 Oct 2017



## must see

a list of 37 titles  
created 1 day ago

The 25 Best War  
Movies of All Time

a list of 25 titles  
created 29 Dec 2016



## Wiederholung - Frage 2

### Frage

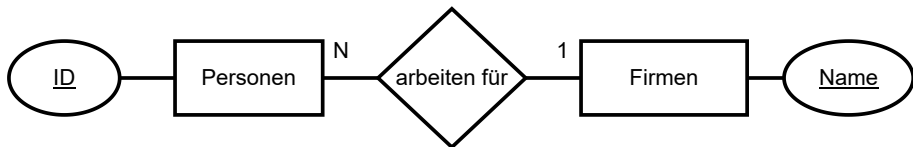
Geben Sie folgendes ER-Modell in vereinfachter Form im relationalen Modell an.



## Wiederholung - Frage 2

### Frage

Geben Sie folgendes ER-Modell in vereinfachter Form im relationalen Modell an.



### Lösung

[Personen] : {[ID:int, Firma\_Name:(Firmen→Name)]}  
[Firmen] : {[Name:string]}

## Wiederholung - Frage 3

### Frage

Mit vielen Autos kann eine Person nach dem unten stehendem Ausschnitt eines Relationenschemas in Beziehung stehen?

$$[\text{sitzen\_in}] : \{ \underbrace{[\text{Person\_ID}:(\text{Personen} \rightarrow \text{ID}), \text{Zeitpunkt}: \text{time}],}_{\text{Auto\_ID}:(\text{Autos} \rightarrow \text{ID})} \}$$

## Wiederholung - Frage 3

### Frage

Mit vielen Autos kann eine Person nach dem unten stehendem Ausschnitt eines Relationenschemas in Beziehung stehen?

$$[\text{sitzen\_in}] : \{ [\text{Person\_ID:}(\text{Personen} \rightarrow \text{ID}), \text{Zeitpunkt: time}, \\ \text{Auto\_ID:}(\text{Autos} \rightarrow \text{ID})] \}$$

### Lösung

Eine Person kann mit beliebig vielen Autos in Beziehung stehen. Lediglich zu einem festen Zeitpunkt kann eine Person in einem einzigen Auto sitzen.

## Wiederholung - Frage 4

Frage

Was ist relationale Algebra?

# Wiederholung - Frage 4

## Frage

Was ist relationale Algebra?

## Lösung

- Anfragesprache zum Kombinieren und Modifizieren von Relationen
- Beschreibt **was** berechnet werden soll, nicht **wie** es berechnet werden soll
- Menge von Operatoren auf Relationen

## Wiederholung - Frage 5

### Frage

Was ist das Ergebnis des folgenden Ausdrucks?

$\sigma_{\text{rank} < 4.0}$  movies

(A):

id	rank
1	2.0

(B):

id	rank
2	4.0

(C):

id	rank
3	3.5

(D):

id	rank
1	2.0
3	3.5

**movies**

id	rank
1	2.0
2	4.0
3	3.5

## Wiederholung - Frage 5

**movies**

id	rank
1	2.0
2	4.0
3	3.5

### Frage

Was ist das Ergebnis des folgenden Ausdrucks?

$\sigma_{\text{rank} < 4.0}$  movies

### Lösung

Die richtige Antwort lautet (D):

id	rank
1	2.0
3	3.5



## Wiederholung - Frage 6

### Frage

Welche Kardinalität (= Anzahl an Tupeln) hat das Ergebnis des folgenden Ausdrucks?

$\pi_{\text{rank}}$  movies

(A): 5

(C): 3

(B): 4

(D): 2

**movies**

id	rank
1	3.5
2	5
3	2.5
4	3.5
5	2.5

## Wiederholung - Frage 6

**movies**

id	rank
1	3.5
2	5
3	2.5
4	3.5
5	2.5

### Frage

Welche Kardinalität hat das Ergebnis des folgenden Ausdrucks?

$\pi_{\text{rank}}$  movies

### Lösung

Die richtige Antwort lautet (C):

Aufgrund der Mengensemantik der relationalen Algebra enthält das Ergebnis der Projektion keine Duplikate und sowohl 2.5, als auch 3.5 werden jeweils nur einmal aufgeführt.

# Wiederholung - Frage 7

## Frage

Was ist das Ergebnis des folgenden Ausdrucks?

movies  $\bowtie$  {[id]}, {[movId]} roles

(A):

id	rank	movId	actId
1	3.5	1	3
1	3.5	1	2
2	4.0		

(B):

id	rank	movId	actId
1	3.5	1	2
1	3.5	1	3

(C):

id	rank	movId	actId
2	4.0	1	2
2	4.0	1	3

(D):

id	rank	movId	actId
1	3.5	1	2
1	3.5	1	3
2	4.0	1	2
2	4.0	1	3

**movies**

id	rank
1	3.5
2	4.0

**roles**

movId	actId
1	2
1	3

## Wiederholung - Frage 7

**movies**

id	rank
1	3.5
2	4.0

**roles**

movId	actId
1	2
1	3

### Frage

Was ist das Ergebnis des folgenden Ausdrucks?

movies  $\bowtie$  {[id]}, {[movId]} roles

### Lösung

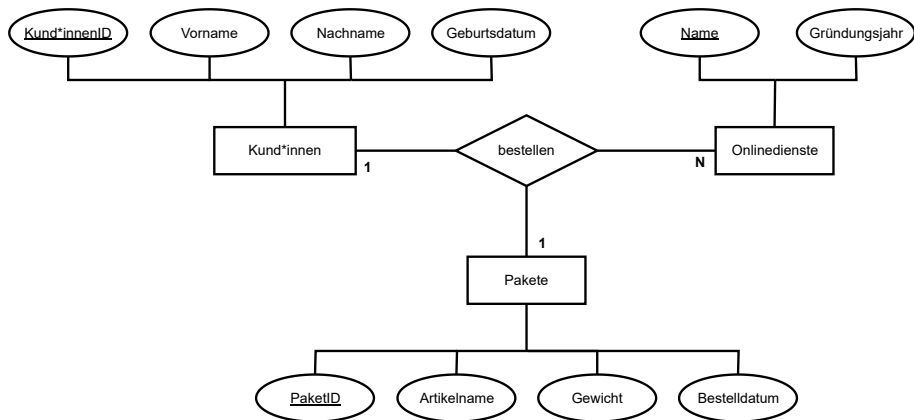
Die richtige Antwort lautet (B):

id	rank	movId	actId
1	3.5	1	2
1	3.5	1	3

# Aufgabe 1.1

## Frage

Welche Schlüsselkombinationen sind bei der Übertragung des Beziehungstypen ins relationale Modell möglich? Stellen Sie für jede Möglichkeit die entsprechende Relation [bestellen] auf.



# Aufgabe 1.1

## Lösung

$$[\text{bestellen}] : \{ [\text{Kund*in:}(\text{Kund*innen} \rightarrow \text{Kund*innenID}), \\ \text{Paket:}(\text{Pakete} \rightarrow \text{PaketID}), \\ \text{Onlinedienst:}(\text{Onlinedienste} \rightarrow \text{Name})] \}$$
$$[\text{bestellen}] : \{ [\text{Kund*in:}(\text{Kund*innen} \rightarrow \text{Kund*innenID}), \\ \text{Paket:}(\text{Pakete} \rightarrow \text{PaketID}), \\ \text{Onlinedienst:}(\text{Onlinedienste} \rightarrow \text{Name})] \}$$

## Aufgabe 1.2

### Frage

Welches Problem ergibt sich durch die Umsetzung ins relationale Modell?  
Welche Information geht durch die Übertragung jeweils verloren?

## Aufgabe 1.2

### Lösung

Je nach Umsetzung geht eine der folgenden Informationen verloren:

- Für ein festes Paket und einen festen Onlinedienst darf nur ein\*e Kund\*in in der Relation sein.
- Für eine\*n feste\*n Kund\*in und einen festen Onlinedienst darf nur ein Paket in der Relation sein.

Beispielsweise wären die Tupel (12, 5, 3) und (12, 2, 3) laut erstem Relationenschema zulässig, widersprechen aber dem zweiten. Umgekehrt wären im zweiten Schema die Tupel (2, 5, 3) und (3, 5, 3) zulässig, nicht aber im ersten.



## Aufgabe 1.3

### Frage

Wie würden Sie das Problem aus Aufgabe 1.2 lösen ohne das relationale Modell zu ändern?

## Aufgabe 1.3

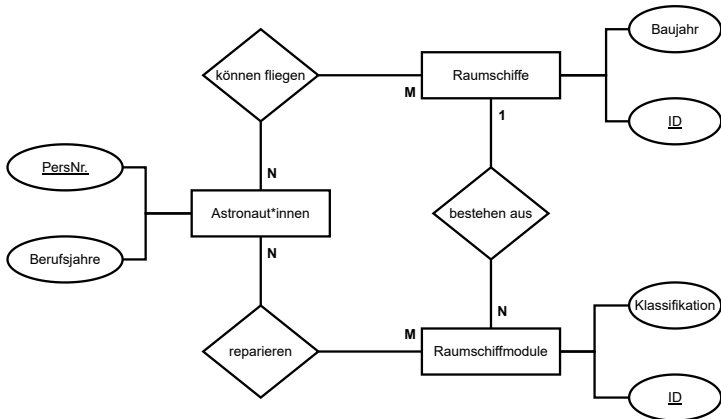
### Lösung

Wir können den Constraint nicht im relationalen Modell darstellen. Da das relationale Modell aber meist nur ein Zwischenschritt zur Tabellendefinition ist, lohnt es sich diese zusätzliche Integritätsbedingung zu vermerken und sie durch entsprechende Maßnahmen bei der Tabellendefinition (beispielsweise UNIQUE) umzusetzen.

## Aufgabe 2

### Frage

Überführen Sie das folgende ER-Diagramm in ein relationales Modell. Vereinfachen Sie das Relationsschema so weit wie möglich.



## Aufgabe 2

### Lösung

$[Astronaut*innen] : \{[\underline{PersNr.:int}, Berufsahre:int]\}$   
 $[Raumschiffe] : \{[\underline{ID:int}, Baujahr:int]\}$   
 $[Raumschiffmodule] : \{[\underline{ID:int}, Raumschiff\_ID:(Raumschiffe \rightarrow ID),$   
 $\quad Klassifikation:string]\}$   
 $[reparieren] : \{[\underline{Astronaut*in\_PersNr.:(Astronaut*innen \rightarrow PersNr.),}$   
 $\quad \underline{Raumschiffmodul\_ID:(Raumschiffmodule \rightarrow ID)}]\}$   
 $[können\ fliegen] : \{[\underline{Astronaut*in\_PersNr.:(Astronaut*innen \rightarrow PersNr.),}$   
 $\quad \underline{Raumschiff\_ID:(Raumschiffe \rightarrow ID)}]\}$

## Aufgabe 3.1

### Frage

Übersetzen Sie folgende umgangssprachliche Anfragen in relationale Algebra:

- (a) Die Wohnorte der Sänger\*innen, die im Rockgenre tätig sind.
- (b) Die Namen der Musiklabels, von denen noch kein Song live gesungen wurde.
- (c) Die Titel der Songs, die live von Sänger\*innen gesungen wurden, die mit ihrem echten Namen auftreten.

## Aufgabe 3.1

### Lösung

- (a)  $\pi_{\text{Wohnort}} ((\sigma_{\text{Genre} = \text{'Rock'}} \text{Sänger*innen}) \bowtie_{\text{SID} = \text{PID}} \text{Personen})$
- (b)  $R_1 := \pi_{\text{Label}}(\text{Songs} \bowtie_{\text{SongID} = \text{Song}} \text{singen\_live})$   
 $R_2 := (\pi_{\text{Label}} (\rho_{\text{Label} \leftarrow \text{MID}} \text{Musiklabels})) - R_1$   
 $\pi_{\text{Name}} (\text{Musiklables} \bowtie_{\text{MID} = \text{Label}} R_2)$
- (c)  $R := \text{Personen} \bowtie_{\text{PID} = \text{SID} \wedge \text{Name} = \text{Künstler*innenname}} \text{Sänger*innen}$   
 $\pi_{\text{Titel}} (\text{Songs} \bowtie_{\text{SongID} = \text{Song}} (\text{singen\_live} \bowtie_{\text{Sänger*in} = \text{SID}} R))$

## Aufgabe 3.2

### Frage

Übersetzen Sie folgende Ausdrücke der relationalen Algebra in umgangssprachliche Anfragen:

- (a)  $\pi_{\text{Genre}} (\text{Sänger*innen} \bowtie_{\text{SID} = \text{Sänger*in}} (\sigma_{\text{Arena} = \text{'SAP'}} \text{singen\_live}))$
- (b)  $R := ((\sigma_{\text{Titel} = \text{'BigDataRocks'}} \text{Songs}) \bowtie_{\text{SongID} = \text{Song}} \text{singen\_live} \pi_{\text{Gründungsjaar}} ((\sigma_{\text{Datum} = 11.05.2007} R) \bowtie_{\text{Label} = \text{MID}} \text{Musiklabels}))$
- (c)  $R := \text{Songs} \bowtie_{\text{SongID} = \text{Song} \wedge \text{Veröffentlichungsdatum} = \text{Datum}} \text{singen\_live} \pi_{\text{Wohnort}} (R \bowtie_{\text{Sänger*in} = \text{PID}} \text{Personen})$

## Aufgabe 3.2

### Lösung

- (a) Die Genres der Sänger\*innen, die schon einmal in der SAP-Arena gesungen haben.
- (b) Das Gründungsjahr der Musiklabels, die schon einmal einen Song mit dem Titel 'BigDataRocks' veröffentlicht haben, und dieser am 11.05.2007 live vorgetragen wurde.
- (c) Der Wohnort der Sänger\*innen, die einen Song an seinem Veröffentlichungsdatum gesungen haben.