Übung zur Vorlesung *Grundlagen: Datenbanken* im WS14/15

Harald Lang (harald.lang@in.tum.de)
http://www-db.in.tum.de/teaching/ws1415/grundlagen/

Blatt Nr. 3

Tool zum Üben der relationalen Algebra `http://www-db.in.tum.de/~muehe/ira/`.

Hausaufgabe 1

Gegeben sei die ER-Modellierung von Zugverbindungen in Abbildung 1. Beachten Sie: `verbindet` modelliert ein **Teilstück** einer Verbindung, d.h. auf der Strecke München → Hamburg gibt es einen Eintrag für die Teilstrecke von München nach Nürnberg, einen Eintrag für Nürnberg nach Würzburg, einen Eintrag für die Teilstrecke Würzburg nach Göttingen und einen Eintrag von Göttingen Nach Hamburg.

a) Übertragen Sie das ER-Modell in ein relationales Schema.

b) Verfeinern Sie das relationale Schema soweit möglich durch Eliminierung von Relationen.

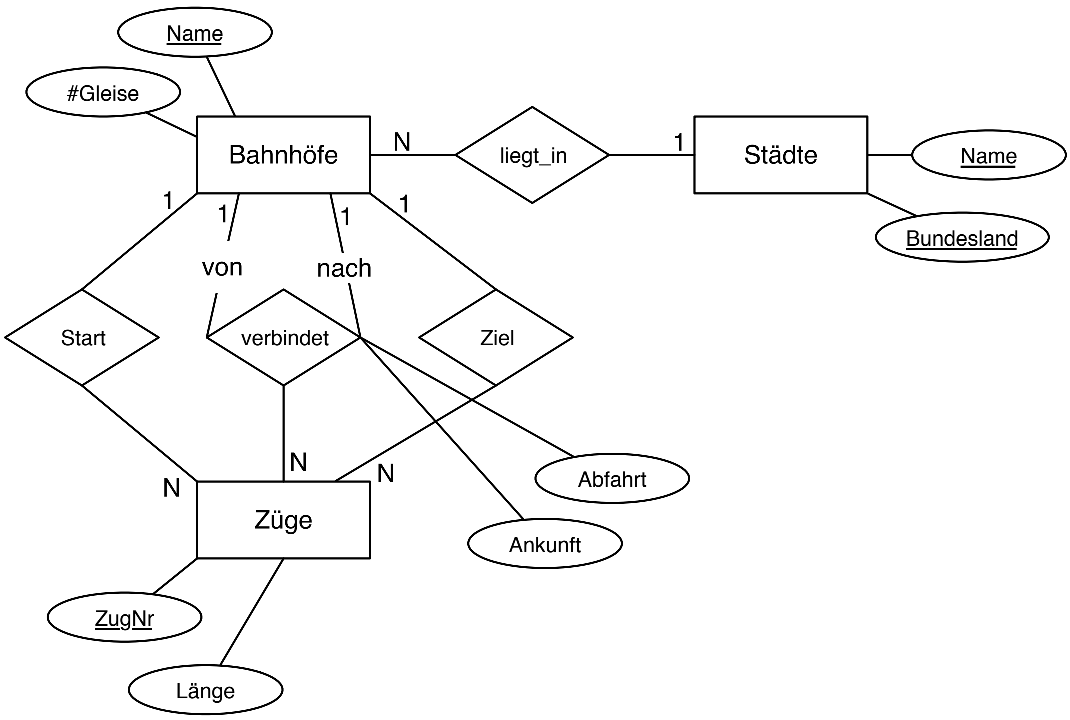


Abbildung 1: ER-Modellierung von Zugverbindungen

Erstellen des relationalen Schemas

Die initiale Überführung ergibt folgende Relationen für die Entitytypen:

$$\begin{aligned}
\text{Städte} &: \{[\underline{\text{Name} : \text{string}, \text{Bundesland} : \text{string}}]\} && (1)\\
\text{Bahnhöfe} &: \{[\underline{\text{Name} : \text{string}}, \#\text{Gleise} : \text{integer}]\} && (2)\\
\text{Züge} &: \{[\underline{\text{ZugNr} : \text{integer}}, \text{Länge} : \text{integer}]\} && (3)
\end{aligned}$$

Für die Beziehungstypen werden folgende Relationen erstellt:

$$\begin{aligned}
\text{liegt_in} &: \{[\underline{\text{BName} : \text{string}}, \text{SName} : \text{string}, \text{Bundesland} : \text{string}]\} && (4)\\
\text{Start} &: \{[\underline{\text{ZugNr} : \text{integer}}, \text{BName} : \text{string}]\} && (5)\\
\text{Ziel} &: \{[\underline{\text{ZugNr} : \text{integer}}, \text{BName} : \text{string}]\} && (6)\\
\text{verbindet} &: \{[\underline{\text{VonBahnhof} : \text{string}}, \text{NachBahnhof} : \text{string}, && (7)\\
&\quad \underline{\text{ZugNr} : \text{integer}}, \text{Abfahrt} : \text{date}, \text{Ankunft} : \text{date}]\}
\end{aligned}$$

Verfeinerung des relationalen Schemas

Als Nächstes wird das relationale Schema verfeinert, indem Relationen zusammengefasst werden.

Dabei werden Relationen für binäre Beziehungstypen mit Relationen für Entitytypen zusammengefasst, falls diese gleiche Schlüssel besitzen und es sich dabei um 1:N, N:1 oder 1:1 Beziehungen handelt.

So kann Relation (4) in (2) aufgenommen werden. (5) wird mit (3) zusammengefasst. Auch die *Ziel*-Relation (6) wird mit der *Züge*-Relation (3) zusammengefasst, d.h.

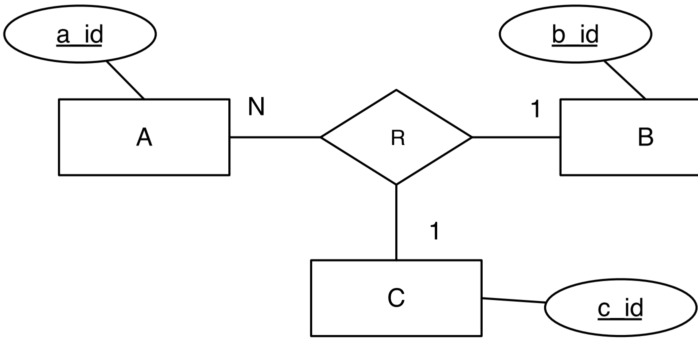
$(4) \mapsto (2)$, $(5) \mapsto (3)$, $(6) \mapsto (3)$

Damit ergibt sich folgendes Schema:

$$\begin{aligned}
\text{Städte} &: \{[\underline{\text{Name} : \text{string}, \text{Bundesland} : \text{string}}]\}\\
\text{Bahnhöfe} &: \{[\underline{\text{Name} : \text{string}}, \#\text{Gleise} : \text{integer},\\
&\quad \text{SName} : \text{string}, \text{Bundesland} : \text{string}]\}\\
\text{Züge} &: \{[\underline{\text{ZugNr} : \text{integer}}, \text{Länge} : \text{integer},\\
&\quad \text{StartBahnhof} : \text{string}, \text{ZielBahnhof} : \text{string}]\}\\
\text{verbindet} &: \{[\underline{\text{VonBahnhof} : \text{string}}, \text{NachBahnhof} : \text{string},\\
&\quad \underline{\text{ZugNr} : \text{integer}}, \text{Abfahrt} : \text{date}, \text{Ankunft} : \text{date}]\}
\end{aligned}$$

Im vorliegenden Fall ist die Zugnummer eindeutig für eine Verbindung. Ein ICE, der die Städte München (*StartBahnhof*) und Berlin (*ZielBahnhof*) verbindet, hat somit eine eindeutige Zugnummer für diese Verbindung, die über mehrere Zwischenbahnhöfe erfolgen kann. Fährt der Zug zurück, erhält er eine andere Nummer zugewiesen. Dadurch sind die Kombinationen (*ZugNr*, *VonBahnhof*) und (*ZugNr*, *NachBahnhof*) zwei mögliche Schlüssel für die Relation *verbindet*.

Hausaufgabe 2



- Welche partiellen Funktionen gelten?
- Setzen Sie das ER Modell in Relationen um.
- Bestimmen Sie einen Schlüssel für die Beziehung `R`, so dass möglichst viele Einschränkungen aus dem ER Modell auch in der Relation für die Beziehung modelliert werden.
- Wieso ist ein Semantikverlust zunächst unvermeidbar? Welche Einschränkung müsste der Relation hinzugefügt werden, um die volle Semantik des ER Modells zu modellieren? [1]

Es gelten die partiellen Funktionen

$$
\begin{aligned}
A \times C &\rightarrow B &\quad (8)\\
A \times B &\rightarrow C. &\quad (9)
\end{aligned}
$$

Aus dem Modell entstehen die folgenden Relationen:

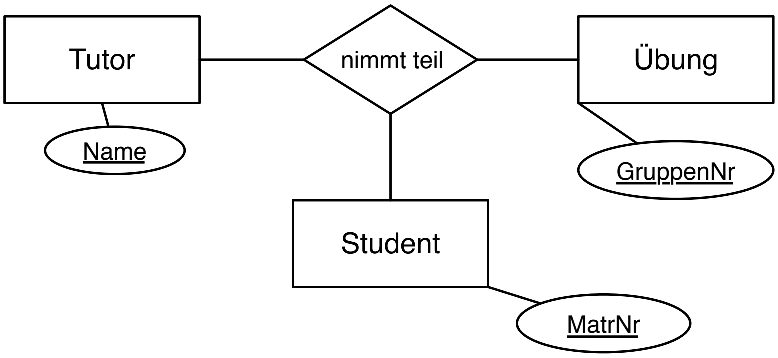
$$
\begin{aligned}
A &: \{[\underline{a\_id}]\} &\quad (10)\\
B &: \{[\underline{b\_id}]\} &\quad (11)\\
C &: \{[\underline{c\_id}]\} &\quad (12)\\
&\text{sowie entweder} & \\
R &: \{[\underline{a\_id}, b\_id, \underline{c\_id}]\} &\quad (13)\\
&\text{oder} & \\
R &: \{[\underline{a\_id}, \underline{b\_id}, c\_id]\} &\quad (14)
\end{aligned}
$$

Egal ob zur Modellierung der Beziehung die Variante in Gleichung 13 oder 14 verwendet wird, kann hier nur die Einschränkung einer der zwei im ER Modell geltenden partiellen Funktionen in die Modellierung mittels Relationen übertragen werden. Eine andere Wahl des Schlüssel außer den hier gezeigten ist entweder falsch, da sie zu starke Anforderungen stellt (etwa wenn lediglich ein Attribut als Schlüssel markiert wurde) oder semantisch schwächer (etwa die Markierung aller Attribute als Schlüssel).

[1] Dieser Teilpunkt ist keine Hausaufgabe. Diskutieren Sie dies in der Übung!

Der Semantikverlust ist zunächst unvermeidbar, da lediglich eine Kombination von Attributen als Primärschlüssel der Beziehung markiert werden kann. Jede Kombination dieser Attribute darf daher nur einmal in der Ausprägung der Beziehungsrelation auftreten, was es erlaubt, die Einschränkung einer der zwei geltenden partiellen Funktionen in die Modellierung als Relation zu übertragen. Die andere geltende partielle Funktion kann nicht in die Relationenmodellierung übernommen werden. Hierfür wäre es notwendig, für eine weitere Menge von Attributen die Einschränkung modellieren zu können, dass jede Kombination dieser Attribute nur einmal in der Ausprägung der Relation auftreten kann. Dies ist beispielsweise in SQL mittels dem `unique`-Schlüsselwort möglich.

Hausaufgabe 3



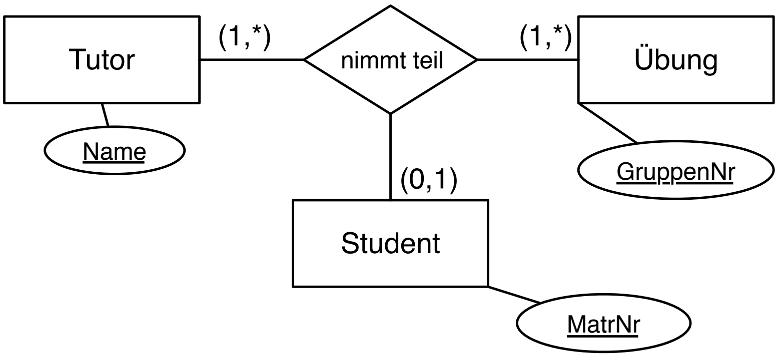
Angenommen, das hier modellierte Übungssystem entspricht dem Übungssystem in Grundlagen: Datenbanken. Bestimmen Sie die MinMax Angaben so, dass folgende Einschränkungen modelliert werden:

- Ein Tutor hält mindestens eine Übung.
- Eine Übung wird von mindestens einem Studenten besucht.
- Ein Student kann höchstens eine Übung besuchen.

Betrachten Sie nun die folgende Ausprägung, die die Beziehung modellieren soll:

Name	GruppenNr	MatrNr
⋮	⋮	⋮
Lang	G12	23
Passing	G27	42
Passing	G27	43
⋮	⋮	⋮
Passing	G28	97
Passing	G28	98
Passing	G28	99
⋮	⋮	⋮

Welche Beziehung besteht zwischen der MinMax Notation und einer solchen Ausprägung?



Im Bezug auf die tabellarische Repräsentation der Beziehung sagen die MinMax Angaben gerade aus, wie oft ein konkreter Wert, etwa der Name des Tutors *Lang*, minimal und maximal vorkommen darf. Die MinMax-Angaben lassen sich also leicht herleiten, indem man sich die tabellarische Repräsentation einer Beziehung vorstellt und überlegt, welche Einschränkungen für die Wiederholung eines konkreten Wertes in dieser Repräsentation gelten sollen. Dadurch, dass im Beispiel keine Matrikelnummer mehrfach vorkommt wird etwa die Einschränkung modelliert, dass ein Student nicht mehrere Übungen besuchen darf.

Hausaufgabe 4

Formulieren Sie folgende Anfragen auf dem bekannten Universitätsschema in der relationalen Algebra, sowie im Tupel- und Domänenkalkül:

a) Finden Sie die *Vorlesungen*, die keine Hörer haben.

b) Finden Sie die *Studenten*, die alle *Vorlesungen* hören.

a) Finden Sie die *Vorlesungen*, die keine Hörer haben.

Formulierung in relationaler Algebra

$$(\ \Pi_{\text{VorlNr}}(\text{Vorlesungen}) \setminus \Pi_{\text{VorlNr}}(\text{hören})\ ) \bowtie \text{Vorlesungen}$$

oder

$$\text{Vorlesungen} \triangleright \text{hören}$$

Formulierung im Tupelkalkül

$$\{\text{v} \mid \text{v} \in \text{Vorlesungen} \wedge \nexists\ \text{h} \in \text{hören}(\text{v.VorlNr} = \text{h.VorlNr})\}$$

oder

$$\{\text{v} \mid \text{v} \in \text{Vorlesungen}\ \wedge \forall\ \text{h} \in \text{hören}(\text{v.VorlNr} \neq \text{h.VorlNr})\}$$

Formulierung im Domänenkalkül

$$\{[\text{v,t,s,g}] \mid [\text{v,t,s,g}] \in \text{Vorlesungen} \wedge \nexists\ \text{m}([\text{m,v}] \in \text{hören})\ \}$$

b) Finden Sie die *Studenten*, die alle *Vorlesungen* hören.

Formulierung in relationaler Algebra

$$(\ \text{hoeren} \div \Pi_{\text{VorlNr}}(\text{Vorlesungen})\ ) \bowtie \text{Studenten}$$

Formulierung im Tupelkalkül

$$\{\text{s} \mid \text{s} \in \text{Studenten}\ \wedge\ \forall\ \text{v} \in \text{Vorlesungen}(\\ \exists\ \text{h} \in \text{hoeren}(\text{v.VorlNr} = \text{h.VorlNr} \wedge \text{s.MatrNr} = \text{h.MatrNr}))\}$$

Formulierung im Domänenkalkül

$$\{[\text{m,n,s}] \mid [\text{m,n,s}] \in \text{Studenten} \wedge \\ \forall\ \text{v,t,sws,g}([\text{v,t,sws,g}] \in \text{Vorlesungen} \Rightarrow [\text{m,v}] \in \text{hören})\}$$

Hausaufgabe 5

Gegeben seien die beiden Relationen $R : \{[a_1, \ldots, a_n]\}$ und $S : \{[b_1, \ldots, b_m]\}$. Geben Sie die folgenden Ausdrücke im Tupel- und Domänenkalkül an:

a) $Q_1 := R \bowtie_{a_1=b_1} S$

b) $Q_2 := R ⟕_{a_1=b_1} S$

c) $Q_3 := R \rtimes_{a_1=b_1} S$

d) $Q_4 := R \lhd_{a_1=b_1} S$

a) $Q_1 := R \bowtie_{a_1=b_1} S$

Formulierung im Tupelkalkül

$$Q_1 := \{[r.a_1, \ldots, r.a_n, s.b_1, \ldots, s.b_m] \mid r \in R\ \wedge\ s \in S\ \wedge\ r.a_1 = s.b_1\}$$

Hinweis: $[r.a_1, \ldots]$ ist eine verkürzte Schreibweise für $[a_1 : r.a_1, \ldots]$

oder noch kürzer:

$$Q_1 := \{r \circ s \mid r \in R\ \wedge\ s \in S\ \wedge\ r.a_1 = s.b_1\}$$

Formulierung im Domänenkalkül

$$Q_1 := \{[a_1, \ldots, a_n, b_1, \ldots, b_m] \mid [a_1, \ldots, a_n] \in R \wedge [b_1, \ldots, b_m] \in S \wedge a_1 = b_1\}$$

oder

$$Q_1 := \{[a_1, \ldots, a_n, b_1 : a_1, b_2, \ldots, b_m] \mid [a_1, \ldots, a_n] \in R \wedge [a_1, b_2, \ldots, b_m] \in S\}$$

b) $Q_2 := R ⟕_{a_1=b_1} S$

Formulierung im Tupelkalkül

$$Q_2 := Q_1 \cup \{[r.a_1, \ldots, r.a_n, b_1 : \text{null}, \ldots, b_m : \text{null}] \mid r \in R \wedge \nexists\, s \in S(r.a_1 = s.b_1)\}$$

Formulierung im Domänenkalkül

$$Q_2 := Q_1 \cup \{[a_1, \ldots, a_n, b_1 : \text{null}, \ldots, b_m : \text{null}] \mid [a_1, \ldots, a_n] \in R \wedge \nexists\, b_2, \ldots, b_m([a_1, b_2, \ldots, b_m] \in S)\}$$

c) $Q_3 := R ⋊_{a_1=b_1} S$

Formulierung im Tupelkalkül

$$Q_3 := \{s \mid s \in S \ \wedge \exists\, r \in R(r.a_1 = s.b_1)\}$$

Formulierung im Domänenkalkül

$$Q_3 := \{[b_1, \ldots, b_m] \mid [b_1, \ldots, b_m] \in S \wedge \exists\, a_2, \ldots, a_n([b_1, a_2, \ldots, a_n] \in R)\}$$

d) $Q_4 := R \triangleleft_{a_1=b_1} S$

Formulierung im Tupelkalkül

$$Q_4 := \{s \mid s \in S \wedge \nexists\, r \in R(r.a_1 = s.b_1)\}$$

Formulierung im Domänenkalkül

$$Q_4 := \{[b_1, \ldots, b_m] \mid [b_1, \ldots, b_m] \in S \wedge \nexists\, a_2, \ldots, a_n([b_1, a_2, \ldots, a_n] \in R)\}$$

Zusatzaufgabe 1 (wird nicht in der Übung besprochen)

Erstellen Sie ein ER-Modell womit sich kausale Zusammenhänge darstellen lassen (Prinzip von Ursache und Wirkung). Nehmen Sie an, dass eine Ursache mehrere Wirkungen haben kann, und dass eine Wirkung auf maximal eine Ursache zurückzuführen ist. Geben Sie die Funktionalitäten an. Verwenden Sie die (min,max)-Notation.

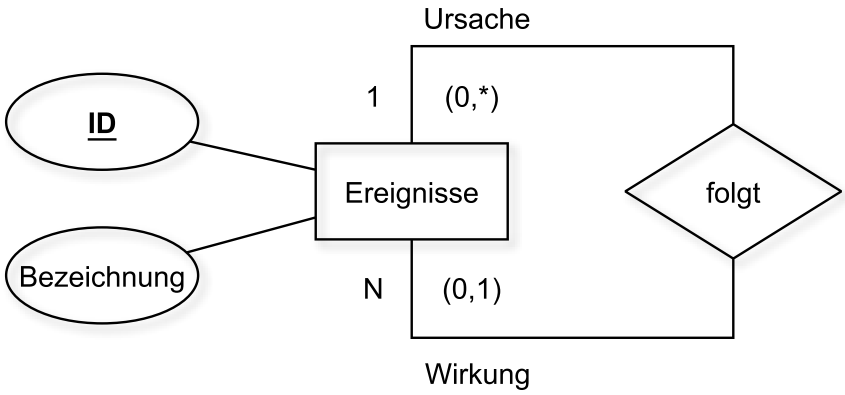


Abbildung 2: ER-Modellierung von kausalen Zusammenhängen

Man kann von Ursache und Wirkung abstrahieren und beides als Ereignis betrachten. Ereignisse können als Ursache sowie als Wirkung auftreten. Für die Relation folgt gilt dann:

$$folgt \subseteq Ursache : Ereignisse \times Wirkung : Ereignisse$$

Eine Ausprägung könnte wie folgt aussehen:

Ursache		*Wirkung*	
ID	*Bezeichnung*	*ID*	*Bezeichnung*
0	Big Bang	1	Beginn der Zeit
	⋮		⋮
10	Fahrzeugpanne	11	Stau auf der A99
11	Stau auf der A99	12	Student verpasst Vorlesung
11	Stau auf der A99	13	Telekom-Techniker verspätet sich mit DSL-Anschluss
	⋮		⋮

(min,max)-Notation

Ursache (0,*)

Sprich: Ein Ereignis kommt in der Rolle "*Ursache*" minimal *0 mal* und maximal *beliebig oft* vor.

Zum aktuellen Zeitpunkt ist noch nicht bekannt, welche Auswirkungen das Ereignis *"Student verpasst Vorlesung"* haben wird, es taucht also 0 mal als Ursache auf. Das Ereignis "*Stau auf der A99*" löst in der oben gezeigten Ausprägung zwei Folgeereignisse aus: "*Student verpasst Vorlesung*" und *"Telekom-Techniker verspätet sich mit DSL-Anschluss*". Denkbar sind beliebig viele (weitere) Auswirkungen.

Wirkung (0,1)

Sprich: Ein Ereignis kommt in der Rolle “*Wirkung*” minimal *0 mal* und maximal *1 mal* vor.

Den Beginn einer Kausalkette bildet ein Ereignis, welches keine Wirkung einer Vorangegangen Ursache ist. So ist der “*Big Bang*” ein Ereignis welches keine (bekannte) Ursache jedoch viele Auswirkungen hat. Er taucht also nur in der Rolle “*Ursache*” auf und somit *0 mal* als Wirkung. Das Ereignis “*Stau auf der A99*” wurde durch die “*Fahrzeugpanne*” verursacht. Nach der Definition in der Aufgabenstellung gilt, dass ein Ereignis maximal eine Ursache hat. Es taucht somit auf der rechten Seite als Wirkung maximal *1 mal* auf.