TU München, Fakultät für Informatik
Lehrstuhl III: Datenbanksysteme
Prof. Alfons Kemper, Ph.D.

Übung zur Vorlesung *Grundlagen: Datenbanken* im WS15/16

Harald Lang, Linnea Passing (gdb@in.tum.de)
http://www-db.in.tum.de/teaching/ws1516/grundlagen/

Blatt Nr. 2

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

Hausaufgabe 1

Formulieren Sie folgende Anfragen auf dem bekannten Universitätsschema in der Relationenalgebra:

(a) Finden Sie die *Assistenten* von *Professoren*, die den Studenten Fichte unterrichtet haben – z.B. als potentielle Betreuer seiner Diplomarbeit.

(b) Finden Sie die *Studenten*, die *Vorlesungen* hören (bzw. gehört haben), für die ihnen die direkten Voraussetzungen fehlen.

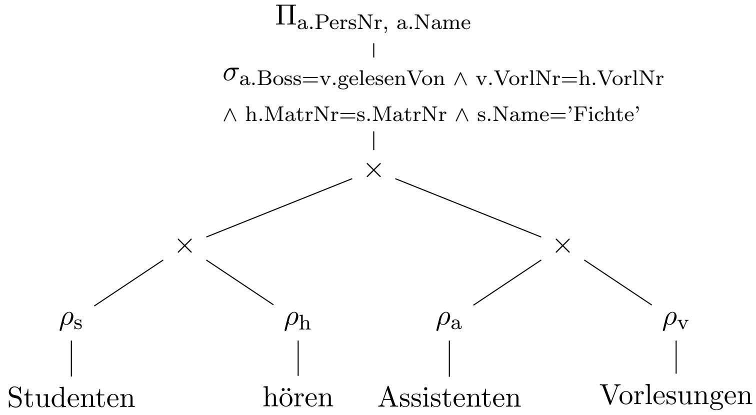
Lösung:

Die Anfragen sehen in relationaler Algebra wie folgt aus:

(a) Folgende Abfrage bildet zuerst das Kreuzprodukt über alle beteiligten Relationen, d.h. *Studenten*, *Vorlesungen*, *Assistenten* und *hören*. Anschließend erfolgt eine umfangreiche Selektion, die die auf Fichte zugeschnittenen Tupel extrahiert.

$$\Pi_{\text{a.PersNr, a.Name}}(\sigma_{\text{a.Boss=v.gelesenVon} \wedge \text{v.VorlNr=h.VorlNr} \wedge \text{h.MatrNr=s.MatrNr} \wedge \text{s.Name='Fichte'}}$$
$$(\rho_{\text{a}}(\text{Assistenten}) \times \rho_{\text{s}}(\text{Studenten}) \times \rho_{\text{v}}(\text{Vorlesungen}) \times \rho_{\text{h}}(\text{hören})))$$

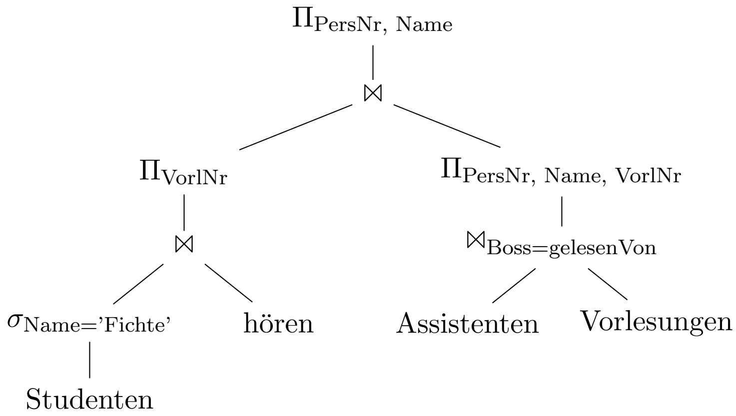
In Operatorbaumdarstellung:



Die Bildung des Kreuzprodukts gilt es nach Möglichkeit zu vermeiden, da dadurch mitunter sehr große Zwischenergebnisse entstehen. Dies kann zu spürbaren Leistungseinbußen während der Anfragebearbeitung führen. Folgende Anfrage berechnet dieselbe Ergebnismenge, setzt jedoch bereits Optimierungstechniken, wie frühe Selektion und den (natürlichen) Verbundoperator ein.

$$\Pi_{\text{PersNr, Name}}((\Pi_{\text{PersNr, Name, VorlNr}}(\text{Assistenten} \bowtie_{\text{Boss=gelesenVon}} \text{Vorlesungen}))$$
$$\bowtie (\Pi_{\text{VorlNr}}(\sigma_{\text{Name='Fichte'}}(\text{Studenten}) \bowtie \text{hören})))$$

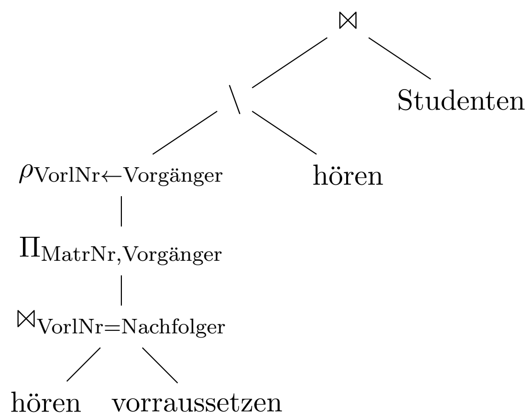
In Operatorbaumdarstellung:



(b) Wir konstruieren eine hypothetische Ausprägung der Relation *hören*, die gelten müsste, wenn alle Studenten alle benötigten Vorgängervorlesungen hören. Von dieser Menge ziehen wir die tatsächliche Ausprägung von hören ab, so dass diejenigen Einträge übrig bleiben, bei denen ein Student die Vorgängervorlesung nicht hört (bzw. gehört hat).

$$\begin{aligned} R \ := \ & (\rho_{\text{VorlNr}\leftarrow\text{Vorgänger}}(\Pi_{\text{MatrNr,Vorgänger}}(\text{hören}\bowtie_{\text{VorlNr=Nachfolger}}\text{voraussetzen})) \\ & -\text{hören})\bowtie\text{Studenten} \end{aligned}$$

In Operatorbaumdarstellung:



Hausaufgabe 2

Gegeben sei die ER-Modellierung von Zugverbindungen in Abbildung 1. Beachten Sie: `verbindet` modelliert ein **Teilstück** einer Verbindung, d.h. auf der Strecke München $\rightarrow$ 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) Fügen Sie bei den Beziehungen Funktionalitätsangaben hinzu.

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

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

Lösung:

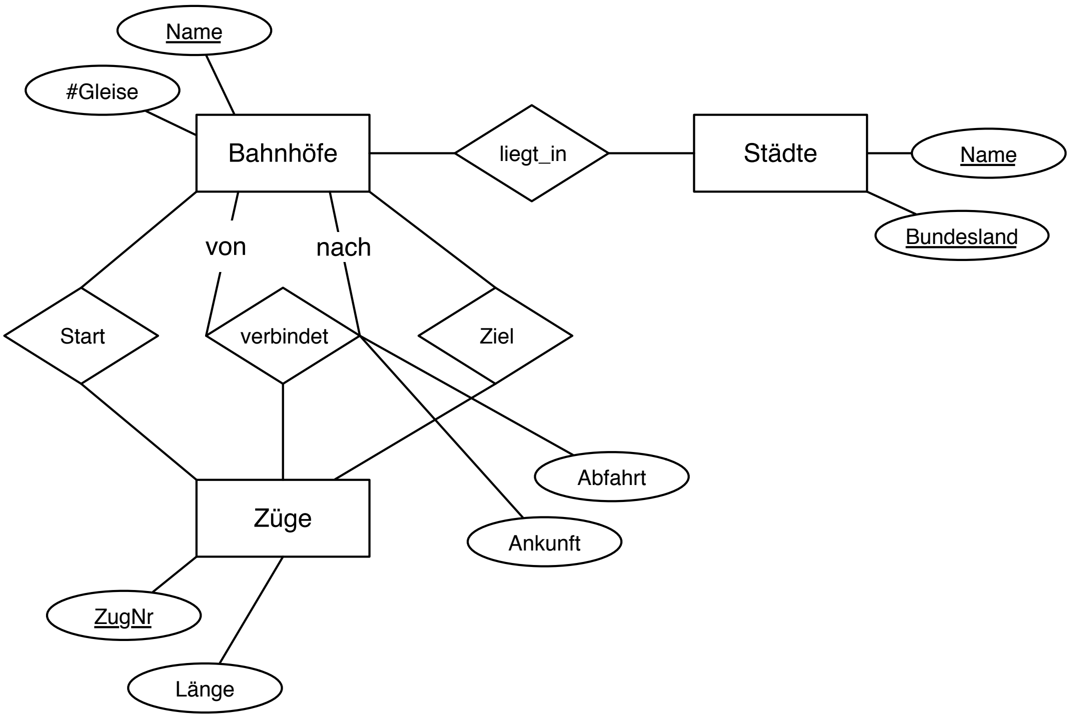


Abbildung 1: ER-Modellierung von Zugverbindungen

a) Eintragen der Multiplizitäten

Abbildung 2 zeigt das ER-Diagramm mit eingetragenen Funktionalitätsangaben.

b) Erstellen des relationalen Schemas

Die initiale Überführung ergibt folgende Relationen für die Entitytypen:

$$\text{Städte} : \{[\underline{\text{Name} : \text{string}, \text{Bundesland} : \text{string}}]\} \tag{1}$$
$$\text{Bahnhöfe} : \{[\underline{\text{Name} : \text{string}}, \#\text{Gleise} : \text{integer}]\} \tag{2}$$
$$\text{Züge} : \{[\underline{\text{ZugNr} : \text{integer}}, \text{Länge} : \text{integer}]\} \tag{3}$$

Für die Beziehungstypen werden folgende Relationen erstellt:

$$\text{liegt\_in} : \{[\underline{\text{BName} : \text{string}}, \text{SName} : \text{string}, \text{Bundesland} : \text{string}]\} \tag{4}$$
$$\text{Start} : \{[\underline{\text{ZugNr} : \text{integer}}, \text{BName} : \text{string}]\} \tag{5}$$
$$\text{Ziel} : \{[\underline{\text{ZugNr} : \text{integer}}, \text{BName} : \text{string}]\} \tag{6}$$
$$\text{verbindet} : \{[\underline{\text{VonBahnhof} : \text{string}}, \text{NachBahnhof} : \text{string}, \underline{\text{ZugNr} : \text{integer}}, \text{Abfahrt} : \text{date}, \text{Ankunft} : \text{date}]\} \tag{7}$$

c) Verfeinerung des relationalen Schemas

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

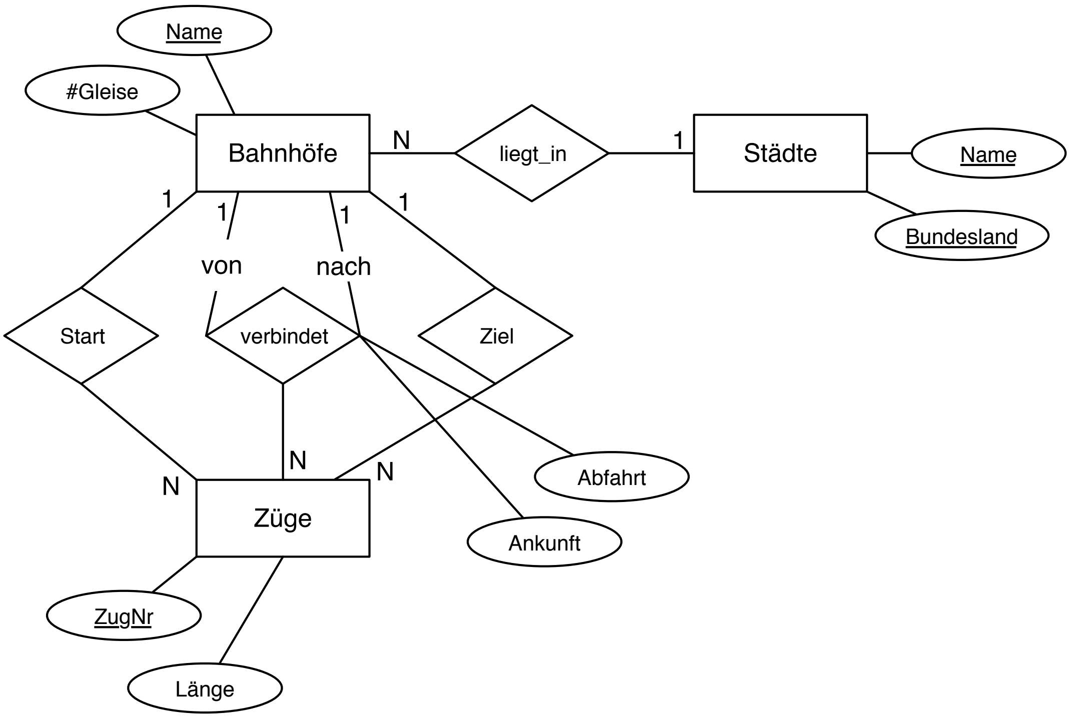


Abbildung 2: ER-Modellierung von Zugverbindungen mit Angabe der Funktionalitäten

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.

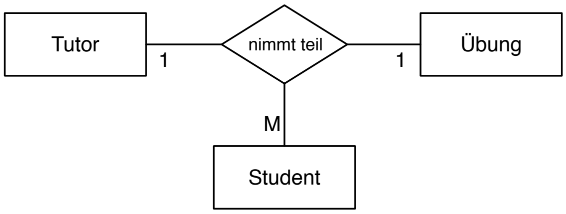
$$(4) \mapsto (2),\ (5) \mapsto (3),\ (6) \mapsto (3)$$

Damit ergibt sich folgendes Schema:

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

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 3



Ignorieren Sie die Funktionalitätsangaben und beantworten Sie:

- Wie viele partielle Funktionen der Form $A \times B \to C$ können in einer ternären Beziehung auftreten (Ignorieren Sie beim Zählen die Reihenfolge auf der linken Seite der Beziehung).
- Nennen Sie alle möglichen partiellen Beziehungen in der hier gezeigten Beziehung „nimmt teil“.
- Nennen Sie für jede Funktion in Prosa, welche Einschränkung diese darstellt, falls sie gilt.

Unter Berücksichtigung der Funktionalitätsangaben:

- Welche partiellen Funktionen gelten hier?

Lösung:

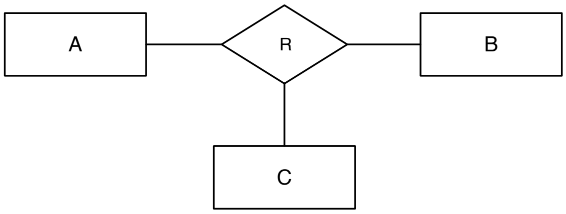
Es gibt drei mögliche partielle Funktionen:

$$Tutor \times Uebung \to Student \tag{8}$$
$$Tutor \times Student \to Uebung \tag{9}$$
$$Uebung \times Student \to Tutor \tag{10}$$

Würde Funktion 8 gelten, so darf ein Tutor pro Übung nur einen Studenten haben. Gilt Funktion 9, so darf ein Student bei einem Tutor nur eine Übung besuchen. Gilt Funktion 10, so darf es für eine konkrete Übung nur einen Tutor geben.

Zu den in der Abbildung gezeigten Kardinalitätsangaben „passen“ die partiellen Funktionen 9 und 10, weshalb diese für das Beispiel gelten. 8 gilt hingegen - wie auch bei uns im Übungsbetrieb - nicht.

Hausaufgabe 4



Angenommen, lediglich die partielle Funktion

$$A \times C \rightarrow B$$

gilt. Beschriften Sie die Abbildung mit Funktionalitätsangaben.

Beantworten Sie nun die Frage, wie Funktionalitätsangaben aus partiellen Funktionen und umgekehrt ermittelt werden können. Merken Sie sich die Antwort für die Klausur ;-)

Lösung:

An der Entität `B` wird eine 1 annotiert, an `A` und `C` jeweils ein N bzw. M.

Eine einfache Daumenregel ist, dass an die Entität, die auf der rechten Seite des Pfeiles einer geltenden partiellen Funktion steht, eine 1 annotiert wird. Es bietet sich daher an, für die sichere Bestimmung der Kardinaliätsangaben grundsätzlich die möglichen partiellen Funktionen aufzustellen und zu überlegen, welche Einschränkungen gewünscht sind.