



Übung zur Vorlesung *Grundlagen: Datenbanken* im WS13/14

Henrik Mühe (muehe@in.tum.de)

<http://www-db.in.tum.de/teaching/ws1314/dbsys/exercises/>

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.

- Übertragen Sie das ER-Modell in ein relationales Schema.
- Verfeinern Sie das relationale Schema soweit möglich durch Eliminierung von Relationen.

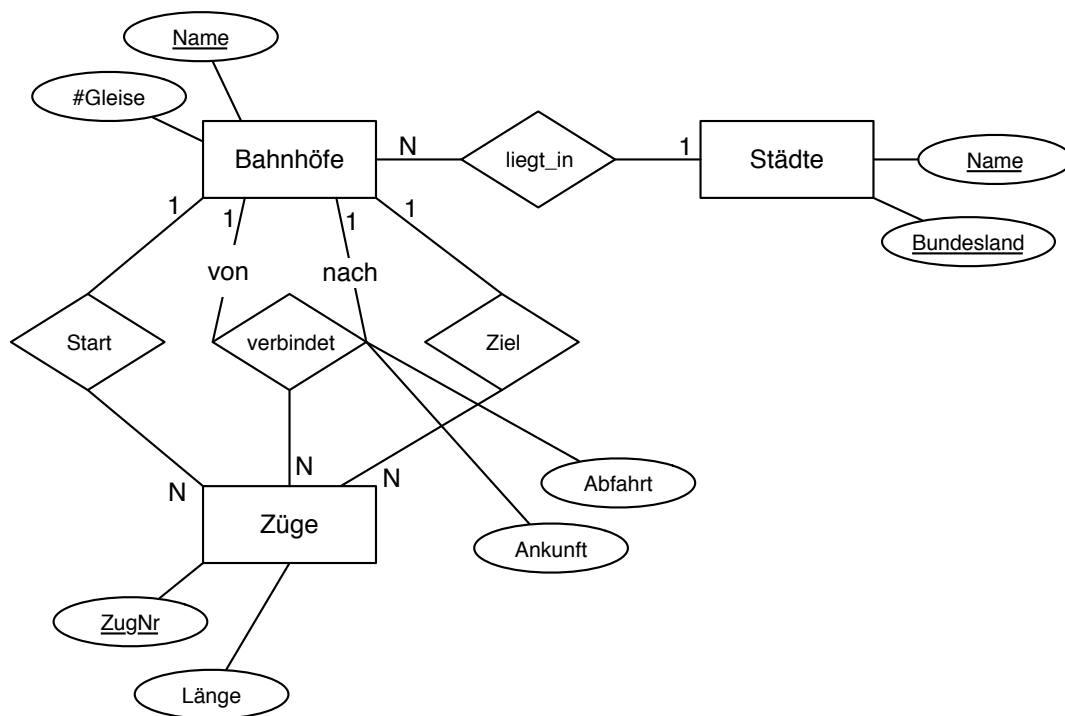


Abbildung 1: ER-Modellierung von Zugverbindungen

Eintragen der Multiplizitäten

Abbildung 2 zeigt das ER-Diagramm mit eingetragenen Funktionalitätsangaben.

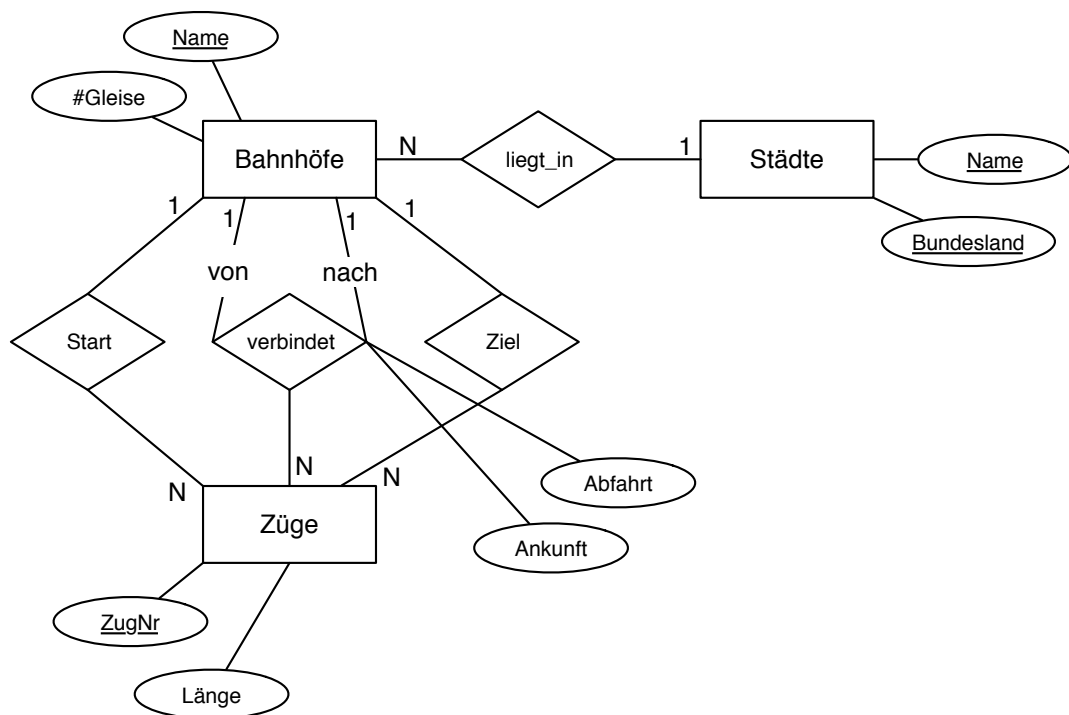


Abbildung 2: ER-Modellierung von Zugverbindungen mit Angabe der Funktionalitäten

Erstellen des relationalen Schemas

Die initiale Überführung ergibt folgende Relationen für die Entitytypen:

$$\text{Städte} : \{[\underline{\text{Name}} : \text{string}, \text{Bundesland} : \text{string}]\} \quad (1)$$

$$\text{Bahnhöfe} : \{[\underline{\text{Name}} : \text{string}, \text{\#Gleise} : \text{integer}]\} \quad (2)$$

$$\text{Züge} : \{[\underline{\text{ZugNr}} : \text{integer}, \text{Länge} : \text{integer}]\} \quad (3)$$

Für die Beziehungstypen werden folgende Relationen erstellt:

$$\text{liegt_in} : \{[\text{BName} : \text{string}, \text{SName} : \text{string}, \text{Bundesland} : \text{string}]\} \quad (4)$$

$$\text{Start} : \{[\underline{\text{ZugNr}} : \text{integer}, \text{BName} : \text{string}]\} \quad (5)$$

$$\text{Ziel} : \{[\underline{\text{ZugNr}} : \text{integer}, \text{BName} : \text{string}]\} \quad (6)$$

$$\text{verbindet} : \{[\underline{\text{VonBahnhof}} : \text{string}, \text{NachBahnhof} : \text{string}, \underline{\text{ZugNr}} : \text{integer}, \text{Abfahrt} : \text{date}, \text{Ankunft} : \text{date}]\} \quad (7)$$

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:

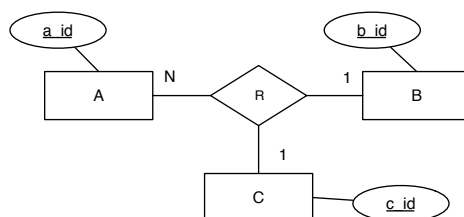
```

Städte : {[Name : string, Bundesland : string]}
Bahnhöfe : {[Name : string, #Gleise : integer,
             SName : string, Bundesland : string]}
Züge : {[ZugNr : integer, Länge : integer,
          StartBahnhof : string, ZielBahnhof : string]}
verbindet : {[VonBahnhof : string, NachBahnhof : string,
              ZugNr : integer, Abfahrt : date, Ankunft : date]}

```

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

$$A \times C \rightarrow B \quad (8)$$

$$A \times B \rightarrow C. \quad (9)$$

Aus dem Modell entstehen die folgenden Relationen:

$$A \quad : \quad \{\underline{a_id}\} \quad (10)$$

$$B \quad : \quad \{\underline{b_id}\} \quad (11)$$

$$C \quad : \quad \{\underline{c_id}\} \quad (12)$$

sowie entweder

$$R \quad : \quad \{\underline{a_id}, \underline{b_id}, \underline{c_id}\} \quad (13)$$

oder

$$R \quad : \quad \{\underline{a_id}, \underline{b_id}, c_id\} \quad (14)$$

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).

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

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

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

Formulieren Sie die folgenden Anfragen auf dem bekannten Universitätsschema in Relationenalgebra.

- (a) Geben Sie alle *Vorlesungen* an, die der *Student* Xenokrates gehört hat.
 - (b) Geben Sie die Titel der direkten Voraussetzungen für die *Vorlesung* Wissenschaftstheorie an.
 - (c) Geben Sie Paare von *Studenten*(-Namen) an, die sich aus der *Vorlesung* Grundzüge kennen.
- (a) Geben Sie alle *Vorlesungen* an, die der *Student* Xenokrates gehört hat.

Formulierung in der Relationenalgebra

$$R := \Pi_{\text{VorlNr, Titel}}(\text{Vorlesungen} \bowtie (\text{hören} \bowtie (\sigma_{\text{Name}='Xenokrates'}(\text{Studenten}))))$$

Formulierung im Tupelkalkül

$$\{v \mid v \in \text{Vorlesungen} \wedge \exists h \in \text{hören}(v.\text{VorlNr} = h.\text{VorlNr} \wedge \exists s \in \text{Studenten}(s.\text{MatrNr} = h.\text{MatrNr} \wedge s.\text{Name} = \text{'Xenokrates'}))\}$$

Formulierung im Domänenkalkül

$$\{[v,t] \mid \exists s,g([v,t,s,g] \in \text{Vorlesungen} \wedge \exists m([m,v] \in \text{hören} \wedge \exists \text{sem}([m,\text{'Xenokrates'},\text{sem}] \in \text{Studenten})))\}$$

- (b) Geben Sie die Titel der direkten Voraussetzungen für die *Vorlesung* Wissenschaftstheorie an.

Formulierung in der Relationenalgebra

$$R := \Pi_{v2.\text{Titel}}(\rho_{v2}(\text{Vorlesungen}) \bowtie_{v2.\text{VorlNr}=\text{Vorgänger}} (\text{voraussetzen} \bowtie_{v1.\text{VorlNr}=\text{Nachfolger}} (\sigma_{v1.\text{Titel}=\text{'Wissenschaftstheorie'}}(\rho_{v1}(\text{Vorlesungen}))))$$

Formulierung im Tupelkalkül

$$\{[v.\text{Titel}] \mid v \in \text{Vorlesungen} \wedge \exists \text{vor} \in \text{voraussetzen}(v.\text{VorlNr} = \text{vor}.\text{Vorgänger} \wedge \exists v2 \in \text{Vorlesungen}(v2.\text{VorlNr} = \text{vor}.\text{Nachfolger} \wedge v2.\text{Titel} = \text{'Wissenschaftstheorie'}))\}$$

Formulierung im Domänenkalkül

$$\{[t] \mid \exists v,s,g([v,t,s,g] \in \text{Vorlesungen} \wedge \exists v2([v,v2] \in \text{voraussetzen} \wedge \exists s2,g2([v2,'Wissenschaftstheorie',s2,g2] \in \text{Vorlesungen})))\}$$

- (c) Geben Sie Paare von *Studenten*(-Namen) an, die sich aus der *Vorlesung* Grundzüge kennen.

Formulierung in der Relationenalgebra

$$\begin{aligned} R := & \Pi_{s1.Name, s2.Name} \\ & (\sigma_{\text{Titel}='Grundzüge'}(\text{Vorlesungen}) \bowtie_{\text{VorlNr}=h1.VorlNr} \\ & (\rho_{s1}(\text{Studenten}) \bowtie_{s1.MatrNr=h1.MatrNr \wedge s1.MatrNr \neq s2.MatrNr} \\ & (\rho_{h1}(\text{hören}) \bowtie_{h1.VorlNr=h2.VorlNr} (\rho_{h2}(\text{hören}))) \\ & \bowtie_{s2.MatrNr=h2.MatrNr} (\rho_{s2}(\text{Studenten})))) \end{aligned}$$

Formulierung im Tupelkalkül

$$\begin{aligned} \{[s1.Name, s2.Name] \mid & s1,s2 \in \text{Studenten} \wedge \exists h1,h2 \in \text{hören} \\ & (s1.MatrNr = h1.MatrNr \wedge s2.MatrNr = h2.MatrNr \wedge \\ & h1.VorlNr = h2.VorlNr \wedge s1.MatrNr \neq s2.MatrNr \wedge \\ & \exists v \in \text{Vorlesungen}(h1.VorlNr = v.VorlNr \wedge v.Titel = 'Grundzüge'))\} \end{aligned}$$

Formulierung im Domänenkalkül

$$\begin{aligned} \{[n1,n2] \mid & \exists m1,m2,s1,s2(m1 \neq m2 \wedge [m1,n1,s1] \in \text{Studenten} \\ & \wedge [m2,n2,s2] \in \text{Studenten} \wedge \exists v([m1,v] \in \text{hören} \\ & \wedge [m2,v] \in \text{hören} \wedge \exists s([v,'Grundzüge',s] \in \text{Vorlesungen})))\} \end{aligned}$$

Hausaufgabe 4

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

- Finden Sie die *Assistenten* von *Professoren*, die den *Studenten* Fichte unterrichtet haben – z.B. als potentielle Betreuer seiner Diplomarbeit.
- Finden Sie die *Studenten*, die *Vorlesungen* hören (bzw. gehört haben), für die ihnen die direkten Voraussetzungen fehlen.

Die Anfragen sehen in relationaler Algebra wie folgt aus:

- 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.

$$\begin{aligned} \Pi_{a.PersNr, a.Name} & (\sigma_{a.Boss=v.gelesenVon \wedge v.VorlNr=h.VorlNr \wedge h.MatrNr=s.MatrNr \wedge s.Name='Fichte'} \\ & (\rho_a(\text{Assistenten}) \times \rho_s(\text{Studenten}) \times \rho_v(\text{Vorlesungen}) \times \rho_h(\text{hören}))) \end{aligned}$$

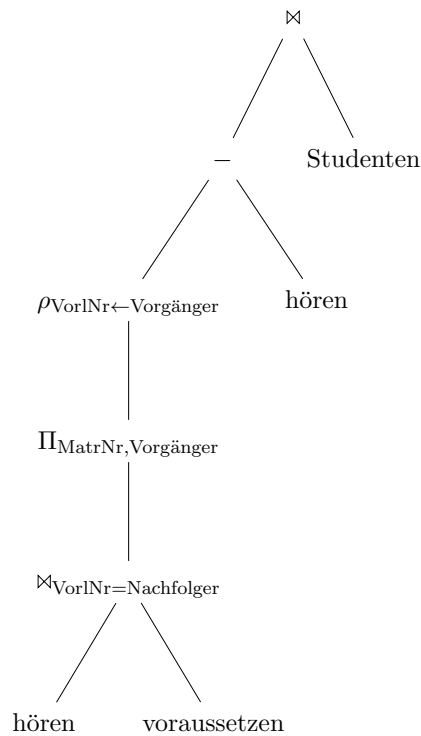


Abbildung 3: Operatorbaum

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.

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

- (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).

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

Abbildung 3 zeigt den zugehörigen Operatorbaum

Gruppenaufgabe 1 (nicht Zuhause vorbereiten)

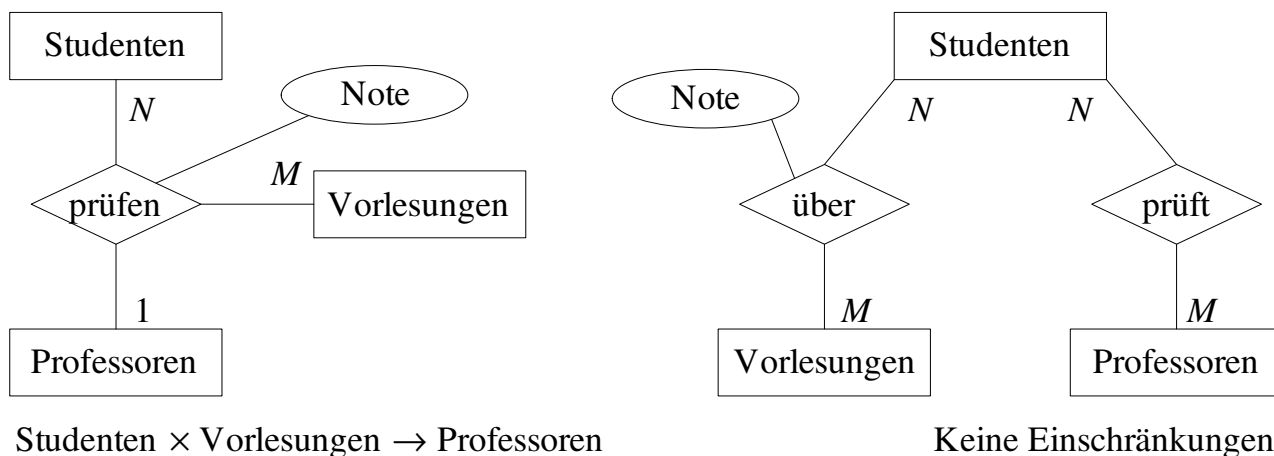


Abbildung 4: Auflösen der ternären Beziehung *prüfen* in binäre Beziehungen

Beim konzeptuellen Entwurf hat man gewisse Freiheitsgrade hinsichtlich der Modellierung der realen Welt. Unter anderem hat man folgende Alternativen, die Sie an unserem Universitätsschema beispielhaft illustrieren sollten:

- Man kann ternäre Beziehungen in binäre Beziehungen transformieren. Betrachten Sie dazu die Beziehung *prüfen* und erläutern Sie die Vor- und Nachteile einer solchen Transformation.
- Man hat manchmal die Wahl, ein Konzept der realen Welt als Beziehung oder als Entitytyp zu modellieren. Erörtern Sie dies wiederum am Beispiel der Beziehung *prüfen* im Gegensatz zu einem eigenständigen Entitytyp *Prüfungen*.
- Ein Konzept der realen Welt kann manchmal als Entitytyp mit zugehörigem Beziehungstyp und manchmal als Attribut dargestellt werden. Ein Beispiel hierfür ist das Attribut *Raum* des Entitytyps *Professoren* im bekannten Uni Schema. Diskutieren Sie die Alternativen.

Ziel dieser Aufgabe ist es, alternative Entwürfe zu erstellen und bezüglich ihrer Anwendbarkeit zu analysieren. Unter Anwendbarkeit ist unter anderem zu verstehen, ob in der neuen Modellierung dieselben Informationseinheiten wie in der ursprünglichen abgebildet werden können, ob Konsistenzbedingungen eingehalten werden und ob die reale Welt in der modellierten Miniwelt sinnvoll wiedergegeben ist.

Erste Teilaufgabe: Transformation der ternären Beziehung in binäre Beziehungen

Abbildung 4 zeigt einen Ansatz, die ternäre Beziehung *prüfen* durch binäre Relationen auszudrücken. Durch die ursprüngliche Modellierung (links in der Abbildung) wird folgende Konsistenzbedingung ausgedrückt:

$$\text{prüfen} : \text{Studenten} \times \text{Vorlesungen} \rightarrow \text{Professoren} \quad (15)$$

Demgegenüber tritt bei der vorgeschlagenen Modellierung mittels binärer Relationen ein Semantikverlust auf. Durch die allgemeineren N:M-Beziehungen wird obige Konsistenzbedingung nicht mehr abgebildet. Somit ist das Modell der ternären Beziehung in diesem Fall ausdrucksstärker. Zwar lassen sich Prüfungsergebnisse in der alternativen Modellierung abbilden, allerdings geht Aussagekraft verloren. Abgebildet ist, dass Studenten über

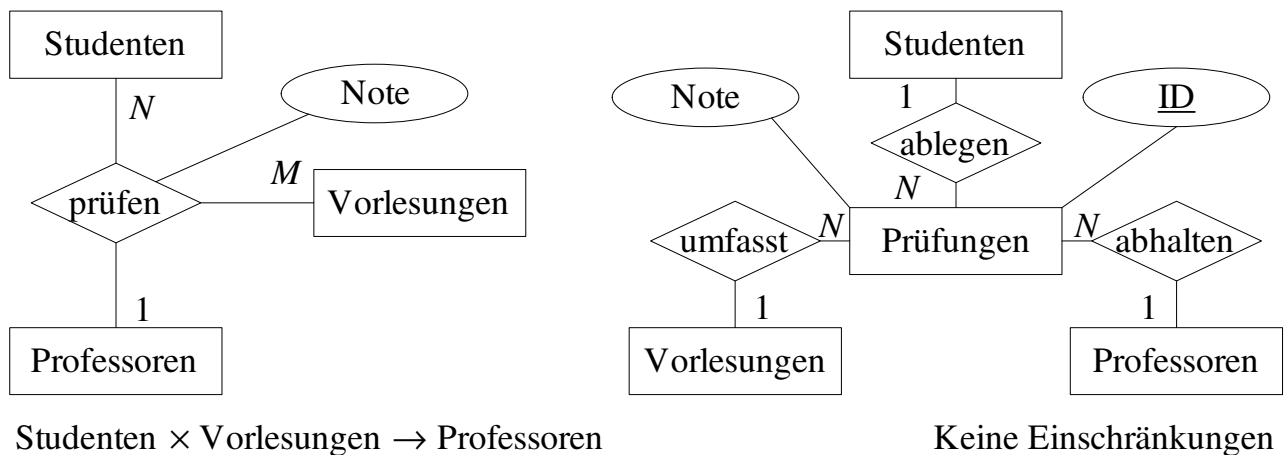


Abbildung 5: Modellierung von Prüfungen als Entitytyp

den Stoff von Vorlesungen geprüft werden, sowie dass Studenten von Professoren geprüft werden. Der Zusammenhang, welche Professoren welche Studenten in welchen Vorlesungen prüfen, ist aber nicht mehr ohne weiteres gegeben. Indirekt lösen lässt sich dies durch die Aufnahme des zusätzlichen Attributs *Prüfungszeit* in die Relation *über* und auch in *prüft*. Da der zusätzlich aufgeführte Prüfungstermin eine Prüfung eindeutig festlegt, lässt sich die Information über eine Prüfung aus beiden Relationen erhalten. Allerdings muss für eine konsistente Extension sichergestellt werden, dass zu einem Eintrag in *über* auch ein passender Eintrag in *prüft* enthalten ist. Die gezeigte alternative Modellierung weist also klare Nachteile gegenüber der ursprünglichen ternären Beziehung auf.

Die alternative Modellierung einer ternären Beziehung durch mehrere binäre kann (abhängig von den zu modellierenden Anforderungen) im Allgemeinen folgende Nachteile aufweisen:

- Es tritt ein Semantikverlust auf.
- Es besteht die Möglichkeit, inkonsistente Datenbankzustände zu generieren. Gegebenenfalls ist eine Konsistenzüberprüfung der Datenbank erforderlich.
- Die reale Welt wird in der Miniwelt unzureichend wiedergegeben.

Zweite Teilaufgabe: Modellierung als Entitytyp anstelle einer Beziehung

Abbildung 5 zeigt eine alternative Modellierung der *prüfen*-Beziehung über einen Entitytyp *Prüfungen*. Auch in diesem Fall tritt erneut ein Semantikverlust auf. Es ist möglich, dass in der Modellierung mittels Entitytyp eine Prüfung existiert, zu der z.B. noch kein Prüfer feststeht, bzw. der Prüfer nicht mehr existiert. Möchte man dies in der Modellierung ausdrücken, müsste man zur *(min,max)*-Notation übergehen, mittels derer man fordern kann, dass eine Prüfung genau je einmal in den Relationen *ablegen*, *abhalten* und *umfasst* auftritt. Außerdem kann auch obige Konsistenzbedingung (15) nicht zugesichert werden. Zwar legt eine *Prüfungen*-Instanz über die angesprochenen Relationen den Studenten / die Studentin, die geprüfte Vorlesung und den / die prüfende(n) Professor(in) fest. Allerdings bestimmt die Kombination *Studenten* \times *Vorlesungen* nun nicht mehr *Professoren*. Denn es ist durch den Entwurf nicht ausgeschlossen, dass es beispielsweise zwei unterschiedliche

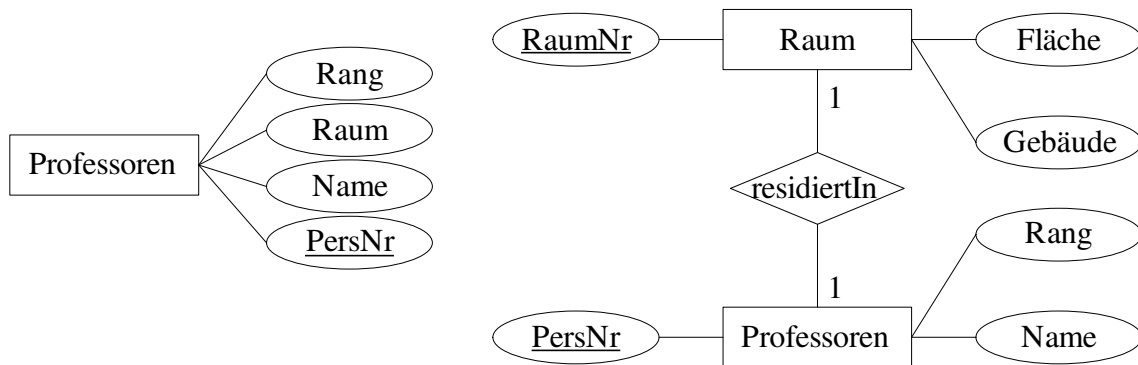


Abbildung 6: Zwei alternative Modellierungen um auszudrücken, dass Professoren ein Raum zugeordnet ist

Prüfungen gibt, die der Student Fichte im Fach Ethik ablegt. Nur einmal lässt er sich bei Professor Sokrates und ein andermal bei Professorin Curie darüber prüfen. Andererseits lassen sich manche Aspekte der Modellierung mittels Entity genauer erfassen, als dies in der ursprünglichen Modellierung der Fall ist. So ist in obigem Beispiel etwa spezifiziert, dass pro Prüfung genau eine Vorlesung geprüft wird.

Dritte Teilaufgabe: Alternative Modellierungen über Attribute oder Beziehungstypen

Abbildung 6 zeigt zwei Modellierungsansätze, die ausdrücken, dass jedem Professor ein Raum zugewiesen ist. Links die Darstellung mittels Attribut, rechts über die Beziehung *residiertIn*. Generell ist eine Modellierung über eine Beziehung mit einem eigenständigen Entity (*Raum*) dann angebracht, wenn entsprechend detaillierte Informationen zu einem Raum nötig sind. Dies kann z.B. dann der Fall sein, wenn die Anwendungssicht der Abteilung Gebäudetechnik in das Modell integriert werden muss. Möchte man die Raumdaten für jeden Professor abfragen, dann zieht diese Modellierung in der Regel eine weniger effiziente Anfrageauswertung nach sich.