



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. 2

Hausaufgabe 1

Unkontrollierte Redundanz ist unerwünscht. Können Sie sich eine sinnvolle Einsatzmöglichkeit für eine durch das DBMS kontrollierte Redundanz vorstellen?

Redundante Datenhaltung zur Performancesteigerung

Kontrollierte redundante Datenhaltung kann aus Performancegesichtspunkten sinnvoll sein. Als Beispiel betrachte man die Personenverwaltung einer Universität. So sind etwa Verwaltungsangestellte, Professoren, wissenschaftliche Mitarbeiter und Studenten Angehörige einer Universität, deren personenbezogene Daten in einem Datenbankverwaltungssystem abgelegt werden. Anfragen, die an die Datenbasis gestellt werden, können sich auf alle Personen beziehen (z.B. Einladung zu einer Universitätsfeier), auf bestimmte Personengruppen (z.B. Erstellen der Lohnabrechnung für alle Angestellten) oder auch nur selektiv auf Teilgruppen (z.B. Einladung zu einer Informationsveranstaltung für alle Studenten im 4. Semester). Die einzelnen Personengruppen sind aber unterschiedlich groß. So ist davon auszugehen, dass es deutlich mehr Studenten als wissenschaftliche Mitarbeiter und mehr wissenschaftliche Mitarbeiter als Professoren gibt. Häufig auftretende Anfragen, die sich nur auf Professoren beziehen, sind dann ineffizient, wenn alle Personendaten zentral in einer Relation abgespeichert sind. In diesem Fall wäre das Überprüfen aller Universitätsangehörigen und das Auswählen der Professoren notwendig.

Speichert man in einer Relation all die Informationen, die sich auf alle Personen beziehen (Name, Anschrift, etc.), und in spezialisierten Relationen die Informationen für Studenten (Name, Anschrift, Semester, Studienrichtung), Professoren (Name, Anschrift, Rang, Raum) usw., so können beide Anfragetypen effizient ausgewertet werden: Es ist dann möglich, allgemeine Anfragen, die sich auf alle Universitätsangehörigen beziehen, wie auch selektive Anfragen, die sich nur auf Personengruppen beziehen, getrennt zu behandeln. Die redundante Datenspeicherung, in diesem Fall also das doppelte Ablegen allgemeiner personenbezogener Daten, wie Name und Anschrift, muss vom Datenbanksystem kontrolliert werden. Um einen konsistenten, d.h. stimmigen Dateninhalt zu gewährleisten, müssen Änderungen des Datenbestands (Einfügen, Löschen oder Modifikation einzelner Datensätze) sowohl in den spezialisierten Datensätzen, wie auch in dem globalen Datensatz der Personendaten durchgeführt werden. Dieses Vorgehen wird auch als Generalisierung bezeichnet.

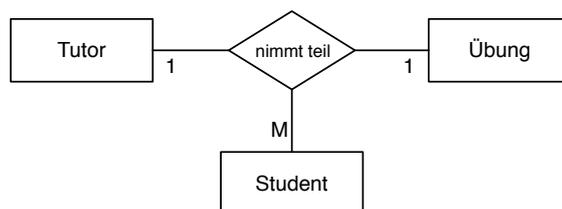
Replizierte Datenhaltung für mobile Anwendungen

Replizierte Datenbanken spielen insbesondere auch bei mobilen Anwendungen eine wichtige Rolle. So werden Teile der globalen Datenbank auf den Laptops von Außendienstmitarbeitern repliziert und regelmäßig (z.B. abends nach den Kundengesprächen) synchronisiert.

Replizierte Datenhaltung bei verteilten Datenbanken

Redundante Datenhaltung ist mitunter auch bei verteilten Informationssystemen erforderlich. Für Anfragen, die von einem verteilten System verarbeitet werden, lässt sich eine höhere Performance ggf. dadurch erreichen, dass häufig abgefragte Daten, die zudem relativ statisch sind, d.h. selten modifiziert werden, an den Knoten im Netzwerk repliziert werden, an denen die Anfrageauswertung erfolgt. Das heißt, die Anfragebearbeitung erfolgt nahe bei den Daten und eine teure Übermittlung von Daten wird so gut es geht vermieden. Je nach Anwendungsfall muss dann natürlich die Konsistenz, d.h. Gültigkeit oder auch "Frische", der replizierten Daten durch das verteilte Datenbanksystem überwacht werden.

Hausaufgabe 2



Ignorieren Sie die Funktionalitätsangaben und beantworten Sie:

- Wie viele partielle Funktionen der Form $A \times B \rightarrow C$ 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?

Es gibt drei mögliche partielle Funktionen:

$$Tutor \times Uebung \rightarrow Student \quad (1)$$

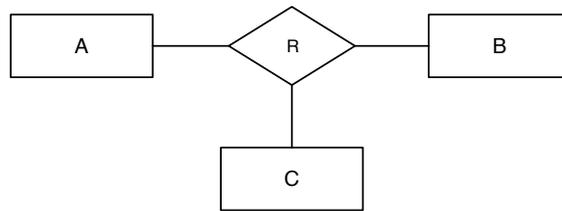
$$Tutor \times Student \rightarrow Uebung \quad (2)$$

$$Uebung \times Student \rightarrow Tutor \quad (3)$$

Würde Funktion 1 gelten, so darf ein Tutor pro Übung nur einen Studenten haben. Gilt Funktion 2, so darf ein Student bei einem Tutor nur eine Übung besuchen. Gilt Funktion 3, so darf es für eine konkrete Übung nur einen Tutor geben.

Zu den in der Abbildung gezeigten Kardinalitätsangaben „passen“ die partiellen Funktionen 2 und 3, weshalb diese für das Beispiel gelten. 1 gilt hingegen - wie auch bei uns im Übungsbetrieb - nicht.

Hausaufgabe 3



Angenommen, lediglich die partielle Funktion

$$A \times C \rightarrow B$$

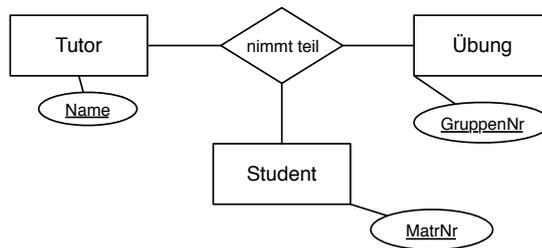
gilt. Beschriften Sie die Abbildung mit Funktionalitätsangaben.

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

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 Kardinalitätsangaben grundsätzlich die möglichen partiellen Funktionen aufzustellen und zu überlegen, welche Einschränkungen gewünscht sind.

Hausaufgabe 4



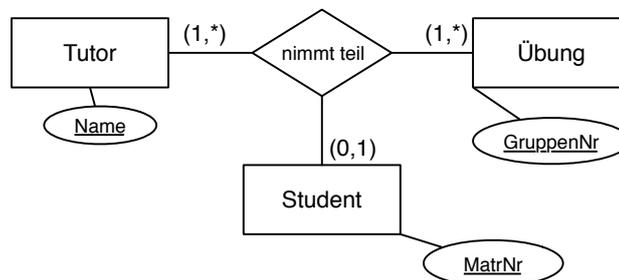
Angenommen, das hier modellierte Übungssystem entspricht dem Übungssystem in Grundlagen: Datenbanken. Bestimmen Sie die MinMax Angaben so, dass folgende Einschränkungen modelliert werden:

- 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
Mühe	G01	23
Merli	G02	42
Merli	G02	43
⋮	⋮	⋮
Merli	G15	97
Merli	G15	98
Merli	G15	99
⋮	⋮	⋮

Welche Beziehung besteht zwischen der MinMax Notation und einer solchen Ausprägung?¹



Im Bezug auf die tabellarische Repräsentation der Beziehung sagen die MinMax Angaben gerade aus, wie oft ein konkreter Wert, etwa der Name des Tutors *Mühe*, minimal und maximal vorkommen darf. Die MinMax-Angaben lassen sich also leicht herleiten, indem

¹Den Spruch zur Klausur spare ich mir an dieser Stelle.

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 5

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) Fügen Sie bei den Beziehungen Funktionalitätsangaben hinzu.

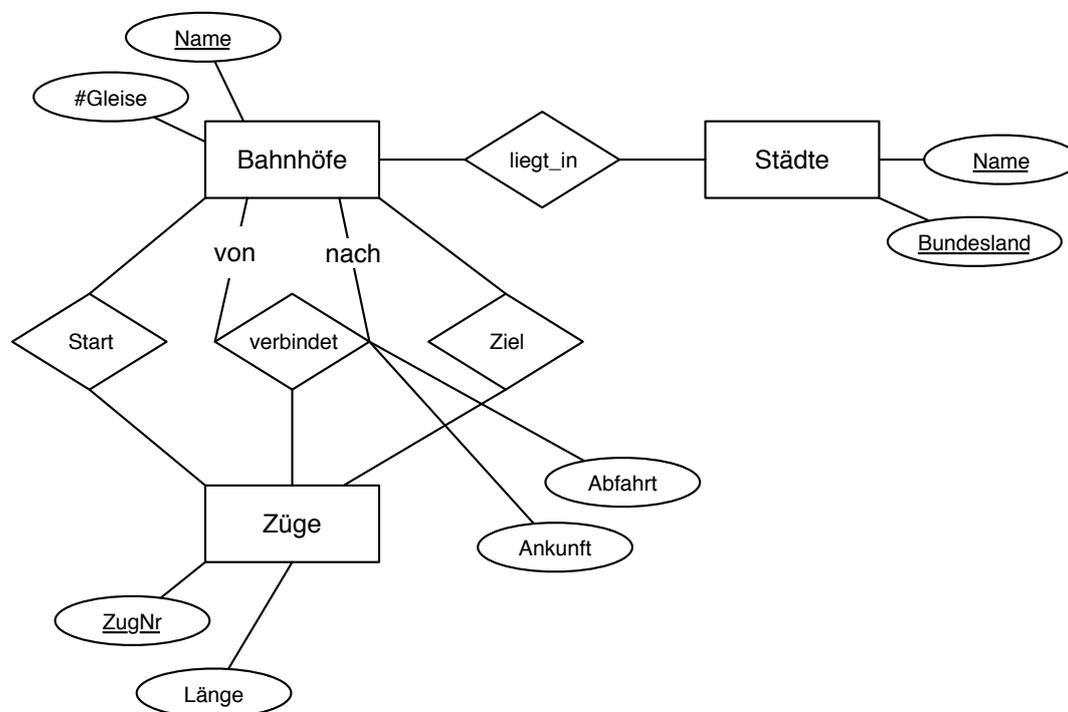


Abbildung 1: ER-Modellierung von Zugverbindungen

Eintragen der Multiplizitäten

Abbildung 2 zeigt das ER-Diagramm mit eingetragenen Funktionalitätsangaben.

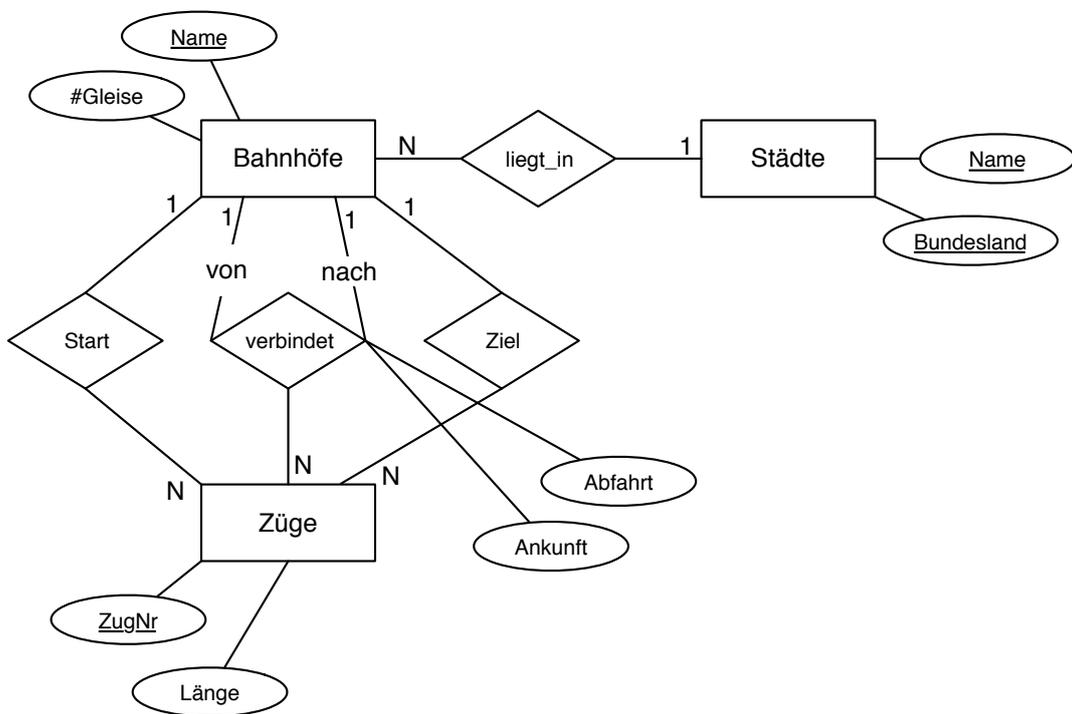


Abbildung 2: ER-Modellierung von Zugverbindungen mit Angabe der Funktionalitäten