

## Übung zur Vorlesung *Grundlagen: Datenbanken* im WS20/21

Christoph Anneser, Josef Schmeißer, Moritz Sichert, Lukas Vogel (gdb@in.tum.de)  
<https://db.in.tum.de/teaching/ws2021/grundlagen/>

### Blatt Nr. 01

Das Uni-Schema finden Sie auf Seite 20 des zweiten Foliensatzes: <https://db.in.tum.de/teaching/bookDBMSeinf/folien/pdf/Kapitel2.pdf>.

### Hausaufgabe 1

Als moderner Netzbürger besitzen Sie einen Twitter-Account mit 150 Followern. Nach der Vorlesung schreiben Sie folgenden Tweet:



- a) Welche Daten werden durch das Absenden dieses Tweets generiert? An welchen Stellen im System von Twitter könnten diese gespeichert werden? Denken Sie nicht nur an den Text des Tweets sondern auch an potentielle Metadaten.

Welche zusätzlichen Daten werden generiert, wenn Ihr Sitznachbar auf das Herz ("Like") klickt?

- b) Seit dem 25.5.2018 gilt die Datenschutz-Grundverordnung (DSGVO), wegen der u.A. personenbezogene Daten auf Anfrage gelöscht werden müssen. Sie sind bei Twitter zuständig, diese Anfragen zu bearbeiten. Überlegen Sie sich, welche Daten bei so einer Anfrage gelöscht werden müssen und welche Probleme dabei auftreten können.

### Lösung:

- a) Natürlich wird der eigentliche Text mit Zuordnung zum Account "@netzbürger42" und dem Datum gespeichert. Da der Tweet eine Referenz auf "@DLR\_de" enthält, wird eine auf diesen Tweet verweisende Benachrichtigung für das DLR erzeugt und gespeichert. Auch durch die Verwendung von Hashtags werden Metadaten erzeugt, um z.B. alle Tweets eines bestimmten Hashtags finden zu können. Alle bisher genannten Daten werden wahrscheinlich in einer relationalen Datenbank bei Twitter in verschiedenen Rechenzentren gespeichert.

Zusätzlich werden indirekte Daten erzeugt. Zum Beispiel kann Twitter Informationen über den Benutzer, wie die IP-Adresse, GPS-Koordinaten, Browser und Betriebssystem, etc. zu statistischen Zwecken speichern. Diese werden wahrscheinlich in einer von den Tweets getrennten Datenbank verwaltet.

Langfristig gesehen macht Twitter auch Backups, in denen die wichtigsten Daten nochmals vorhanden sind. Außerdem könnte Twitter auch Informationen zu Tweets und Hashtags weiteren Unternehmen zur Verfügung stellen, die diese dann weiter auswerten, z.B. zu Werbezwecken.

Wenn Ihr Nachbar auf das Herz klickt, wird eine Verbindung zwischen dessen Account und Ihrem Tweet gespeichert.

- b) Zunächst einmal müssen alle Tweets sowie das Nutzerkonto selbst gelöscht werden. Das umfasst Daten wie Account-Name, Password-Hashes, Anschrift, Geburtstag, etc. Da diese wahrscheinlich an einer Stelle gespeichert werden, sind sie einfach zu löschen.

Komplizierter wird es in den Fällen, wo der zu löschende Nutzer mit anderen Nutzern interagiert hat. Hier ist es wichtig, sich ein Vorgehen zu überlegen, dass die Konsistenz aller Daten garantiert. Was zum Beispiel geschieht mit Direktnachrichten? Werden nur die Nachrichten des Nutzers gelöscht, sodass der Gesprächspartner nur noch seine Antworten sieht? Oder wird das gesamte Gespräch gelöscht? Auch denkbar wäre es, die Nachrichten zu erhalten, aber den Namen durch ein Pseudonym zu ersetzen (so macht es Facebook). Dasselbe Problem existiert für Antworten auf Tweets oder Retweets.

Noch schwieriger wird es mit Daten, auf welche nicht direkt zugegriffen werden kann. Das trifft insbesondere auf Backups zu, welche oft komprimiert, oder auf Offline-Speichersystemen (Band) bzw. Read-Only-Medien (Optische Medien) gespeichert sind. Twitter kann nicht wegen jedem Löschantrag seine gesamten Backups zerstören. Twitter könnte hier z.B. Backups verschlüsseln, sodass "gelöschte" Daten nicht in die falschen Hände gelangen können sowie ihr Recovery-System so anpassen, dass es Daten von "gelöschten" Nutzern nicht wiederherstellt, sollte eine Systemwiederherstellung nötig sein.

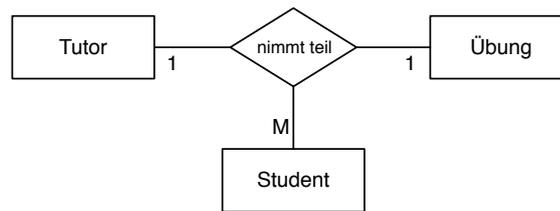
## Hausaufgabe 2

Finden Sie ein Beispiel für ein Problem (bzw. eine Inkonsistenz), die auftreten kann, wenn unkontrolliert parallel auf Daten zugegriffen wird. Ein traditionelles Beispiel hierfür ist eine gegenseitige Bank-Überweisung zwischen zwei Konten A und B. Wenn A einen Betrag  $x$  zu B überweist und B einen Betrag  $x'$  zu A, sollte immer gelten  $Kontostand(A) + Kontostand(B)$  ist konstant, da sonst Geld verschwunden ist. Konstruieren Sie einen Ablauf zweier gegenseitiger Überweisungen, bei dem die Eigenschaft, dass die Kontostandssumme konstant sein soll nach dem Abschluss der zwei Überweisungen verletzt ist.

### Lösung:

- A liest eigenen Kontostand in Variable  $a$  ein.
- A dekrementiert  $a$  um  $x$ .
- B liest eigenen Kontostand in Variable  $b$  ein.
- B dekrementiert  $b$  um  $x'$
- B liest As Kontostand in Variable  $a'$  ein.
- B inkrementiert  $a'$  um  $x'$ .
- B schreibt  $a'$  in As Kontostand zurück. ...

### Hausaufgabe 3



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?

**Lösung:**

- Es gibt drei mögliche partielle Funktionen
- 

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

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

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

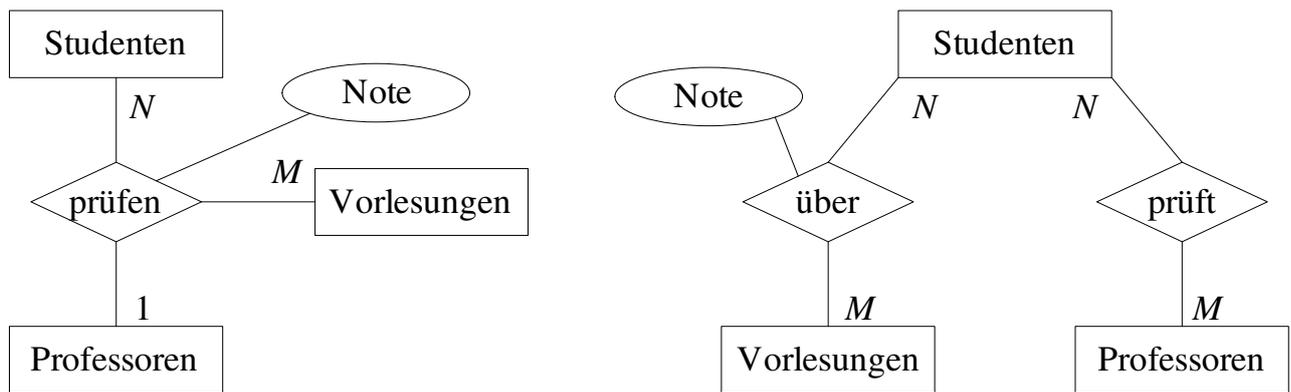
- Würde Funktion 1 gelten, so darf ein Tutor pro Übung nur einen Studenten haben. Gilt Funktion 2, so darf ein Student bei einem Tutor nur eine Übung besuchen. Gilt Funktion 3, so darf es für einen konkreten Studenten in einer Übung nur einen Tutor geben.
- Zu den in der Abbildung gezeigten Kardinalitätsangaben „passen“ die partiellen Funktionen 2 und 3, weshalb diese für das Beispiel gelten. 1 gilt hingegen - wie auch bei uns im Übungsbetrieb - nicht.

### Hausaufgabe 4

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.



Studenten  $\times$  Vorlesungen  $\rightarrow$  Professoren

Keine Einschränkungen

Abbildung 1: Auflösen der ternären Beziehung *prüfen* in binäre Beziehungen

- 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.

#### Lösung:

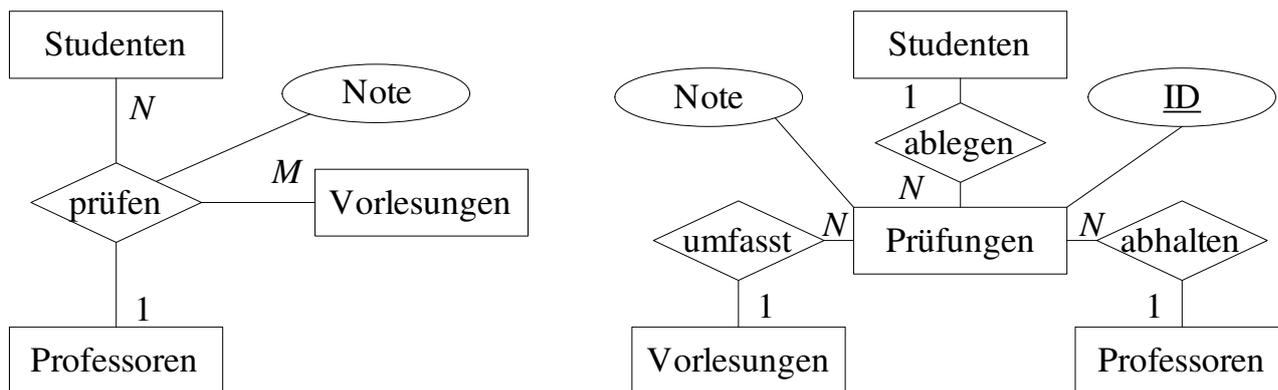
Ziel dieser Aufgabe ist es, alternative Entwürfe zu erstellen und bezüglich ihrer Anwendbarkeit zu analysieren. Unter Anwendbarkeit ist unter anderem zu verstehen, ob in der neuen Modellierung dieselben Informationseinheiten wie in der ursprünglichen abgebildet werden können, ob Konsistenzbedingungen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

#### Erste Teilaufgabe: Transformation der ternären Beziehung in binäre Beziehungen

Abbildung 1 zeigt einen Ansatz, die ternäre Beziehung *prüfen* durch binäre Relationen auszudrücken. Durch die ursprüngliche Modellierung (links in der Abbildung) wird folgende Konsistenzbedingung ausgedrückt:

$$\text{prüfen} : \text{Studenten} \times \text{Vorlesungen} \rightarrow \text{Professoren} \quad (4)$$

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 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 *Prüfungstermin* in die Relation *über* und auch in *prüft*. Da der zusätzlich aufgeführte Prüfungstermin eine Prüfung eindeutig festlegt, lässt sich die



Studenten  $\times$  Vorlesungen  $\rightarrow$  Professoren

Keine Einschränkungen

Abbildung 2: Modellierung von Prüfungen als Entitytyp

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 2 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 (4) 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 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.

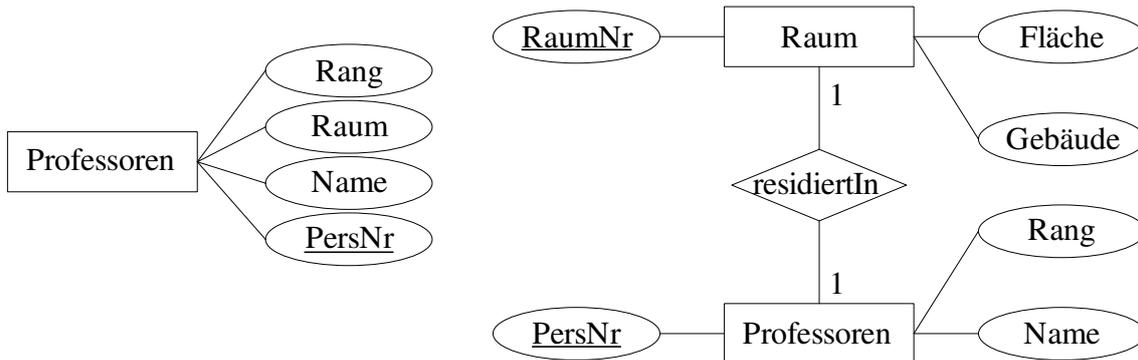


Abbildung 3: Zwei alternative Modellierungen um auszudrücken, dass Professoren ein Raum zugeordnet ist

### Dritte Teilaufgabe: Alternative Modellierungen über Attribute oder Beziehungstypen

Abbildung 3 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.