

1. Übungsblatt

3.0 VU Datenmodellierung

12. November 2014

Allgemeines

In diesem Übungsteil werden Sie den Entwurf einer kleinen Datenbank üben, deren Überführung in das Relationenschema, sowie die relationale Algebra und den Relationenkalkül kennenlernen.

Lösen Sie die Beispiele **eigenständig**, denn bei der Prüfung und in der Praxis sind Sie auch auf sich alleine gestellt. Wir weisen Sie darauf hin, dass abgeschriebene Lösungen mit 0 Punkten beurteilt werden.

Geben Sie ein einziges PDF Dokument ab. Erstellen Sie Ihr Abgabedokument computerunterstützt. Wir akzeptieren keine gescannten handschriftlichen PDF-Dateien.

Deadlines

28.11.	06:55 Uhr	Upload über den COURSEMANAGER
06.12.	08:00 Uhr	Feedback im COURSEMANAGER verfügbar
07.12.	23:59 Uhr	Reservierung eines Termins für das Tutorengespräch
07.12.	23:59 Uhr	Reservierung eines Termins für das Abgabegespräch

Tutorengespräch

1. Sie müssen sich über den COURSEMANAGER zu einem Tutorengespräch anmelden. Bitte machen Sie das rechtzeitig, je später Sie sich anmelden, umso eingeschränkter ist das Terminangebot.
2. Wenn Sie kein Blatt abgegeben haben, sind Sie nicht zum Tutorengespräch zugelassen.
3. Sie kommen mit Ihrem Studierendenausweis zu der von Ihnen reservierten Zeit vorbei, und absolvieren das Tutorengespräch. Sie erhalten dabei wertvolles Feedback zu Ihrem Übungsblatt. Das Gespräch ist verpflichtend und dauert ca. 30 Minuten.

Abgabegespräch

1. Sie müssen sich über den COURSEMANAGER zu einem Abgabegespräch anmelden. Bitte machen Sie das rechtzeitig, je später Sie sich anmelden, umso eingeschränkter ist das Terminangebot.
2. Wenn Sie kein Blatt abgegeben haben oder kein Tutorengespräch geführt haben, sind Sie nicht zum Abgabegespräch zugelassen.
3. Sie kommen mit Ihrem Studierendenausweis zu der von Ihnen reservierten Zeit vorbei, und absolvieren das Abgabegespräch. Stoffgebiet des Abgabegesprächs sind die mit dem Übungsblatt abgedeckten Themengebiete. Wir setzen voraus, dass Sie sich mit Ihrer korrigierten Abgabe auseinandergesetzt haben. Das Abgabegespräch hat Prüfungscharakter.
4. Sie absolvieren Ihr Abgabegespräch gemeinsam mit anderen KollegInnen. Das Gespräch dauert ca. 60 Minuten.
5. Sie können auf die Abgabe maximal 15 Punkte erreichen. Diese setzen sich wie folgt zusammen:
5 Punkte auf das Übungsblatt
10 Punkte auf das Abgabegespräch
6. Die Assistenten tragen die Punkte des Abgabegesprächs in den COURSEMANAGER ein und Sie sehen dort, wieviele Punkte Sie bekommen haben.
7. Falls Sie nicht zu Ihrem Gesprächstermin erscheinen, bekommen Sie automatisch 0 Punkte auf das Gespräch.

EER

Aufgabe 1 (EER-Diagramm erstellen) [0.5 Punkte]

Zeichnen Sie folgenden Sachverhalt in einem EER-Diagramm. Verwenden Sie dazu die (min, max)-Notation. Nullwerte sind nicht erlaubt. Manchmal kann es notwendig sein, zusätzliche künstliche Schlüssel einzuführen. Eine Unterstützung bei der Erstellung von EER-Diagrammen bietet das Tool dia (<http://live.gnome.org/Dia>). Achtung: im Diagramm Editor EER auswählen!

Die Geschäftsführung des Literaturverlags Leserate möchte eine Datenbank anlegen, um die wichtigsten Informationen über die Veröffentlichungen des Verlags zu speichern.

In der Datenbank soll gespeichert werden, welche Personen an den Veröffentlichungen mitarbeiten. Personen werden durch eine eindeutige ID (PID) identifiziert. Außerdem wird zu jeder Person der Name (NAME), das Geburtsdatum (GEBDAT) und eine Kurzbiographie für die Website des Verlags (KURZBIO) gespeichert.

Bei den Personen kann es sich um Autor/innen, Übersetzer/innen oder Lektor/innen handeln. Eine Person kann auch Autor/in und Übersetzer/in zugleich sein.

Bei Lektor/innen wird in der Datenbank erfasst, welche Sprachen sie beherrschen. Jede Sprache wird dabei durch einen eindeutigen Namen (NAME) identifiziert. Ein Lektor oder eine Lektorin kann beliebig viele Sprachen beherrschen. Bei Übersetzer/innen muss gespeichert werden, in welche Sprachen sie übersetzen können und aus welchen Originalsprachen sie übersetzen können. Es kann sein, dass jemand z.B. englische Texte ins Deutsche übersetzen kann, aber keine deutschen Texte ins Englische.

Für jedes Werk, das im Verlag Leseratte erscheint, wird gespeichert, welche Autor/innen es geschrieben haben. Dabei ist zu beachten, dass auch mehrere Autor/innen gemeinsam ein Werk schreiben können. Ein Werk wird durch den Titel (TITEL) und das Jahr, in dem es geschrieben wurde (JAHR), identifiziert. Es wird auch eine kurze Beschreibung für die Verlagswebsite (BESCHREIBUNG) gespeichert.

Ein Werk kann eine Übersetzung eines anderen Werks sein. In der Datenbank soll gespeichert werden, bei welchen Werken es sich um Übersetzungen handelt und wer sie übersetzt hat. Der Einfachheit halber nehmen wir an, dass es nicht möglich ist, dass mehrere Personen gemeinsam eine Übersetzung eines Werks erstellen. Von einem Werk kann es aber mehrere Übersetzungen geben, die auch von unterschiedlichen Übersetzer/innen sein können.

Zu den Werken wird außerdem noch gespeichert, zu welchen Genres (Science-Fiction-Roman, Gedichtband, ...) sie gehören. Jedes Werk wird einem oder mehreren Genres zugeordnet. Jedes Genre wird durch eine ID (GID) eindeutig identifiziert und hat außerdem noch einen Namen (NAME) und eine kurze Beschreibung (BESCHREIBUNG). Ein Genre kann (muss aber nicht) ein Subgenre von einem oder mehreren anderen Genres sein.

Von jedem Werk, das im Verlag Leseratte erscheint, gibt es eine oder mehrere Auflagen. Eine Auflage wird durch das Werk und durch eine zusätzliche Auflagennummer (NUMMER) eindeutig identifiziert. (Hinweis: Die Auflagennummer allein ist im Allgemeinen nicht eindeutig.) Außerdem soll das Jahr gespeichert werden, in dem die Auflage erscheint (JAHR), und die Lektor/innen, die an der Auflage mitgearbeitet haben. An jeder Auflage arbeitet mindestens ein Lektor/eine Lektorin mit.

Zu einer Auflage eines Werks kann es mehrere Publikationen geben. Publikationen können die Typen Taschenbuch, Hardcover und E-Book haben, wobei es zu einem Werk auch mehrere Publikationen vom selben Typ geben kann. Publikationen werden durch die Auflage und eine zusätzliche ID (PID) identifiziert. Außerdem wird noch die Seitenanzahl (SEITEN) und bei E-Books das Dateiformat (FORMAT) gespeichert. Bei manchen, aber nicht allen Publikationen gibt es einen Klappentext (TEXT), der auch in der Datenbank

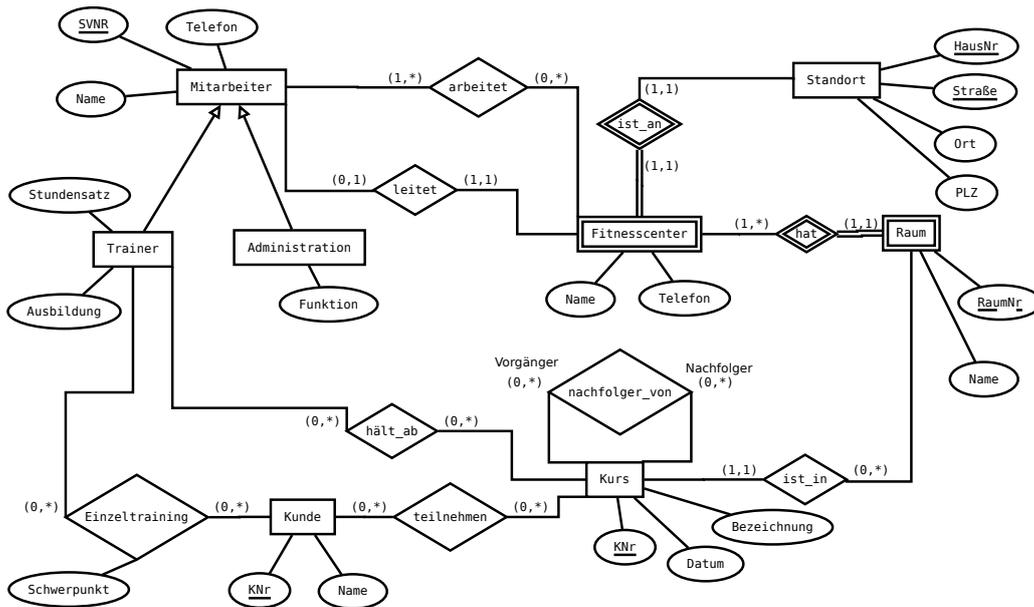


Abbildung 1: EER-Diagramm zu Aufgabe 3

gespeichert werden soll.

Schließlich sollen auch noch die Verträge, die der Verlag mit Personen abschließt, in die Datenbank aufgenommen werden. Zu jedem Vertrag enthält die Datenbank eine eindeutige ID (VID), das Abschlussdatum (DATUM), einen kurzen Beschreibungstext (BESCHREIBUNG) und – optional – eine Datei mit dem Vertragstext (DATEI). Außerdem soll gespeichert werden, mit welcher Person der Vertrag abgeschlossen wird. Dabei kann man davon ausgehen, dass Verträge immer mit genau einer Person abgeschlossen werden.

Aufgabe 2 (EER-Beispiel selbst erstellen) [0.5 Punkte]

Ihr Kollege Herr Mustermann hat Sie gebeten, ihm ein wenig Nachhilfe zu geben. Erstellen Sie für ihn ein kleines Übungsbeispiel zum Thema EER-Modellierung. Das Beispiel sollte mindestens 5 Entitäten, 5 Beziehungen, eine Generalisierung und eine schwache Entität enthalten. Vielleicht bringen Sie in dem Beispiel auch noch eine rekursive oder eine ternäre Beziehung unter. Sie erhalten die Punkte für die textuell ausformulierte Angabe sowie für die Musterlösung.

Aufgabe 3 (Überführung ins Relationenschema) [0.5 Punkte]

Überführen Sie das EER-Diagramm aus Abbildung 1 ins Relationenschema. Nullwerte sind nicht erlaubt. Verwenden Sie möglichst wenig Relationen. Unterstreichen Sie sämtliche Primärschlüssel, schreiben Sie die Fremdschlüssel kursiv und stellen Sie sicher, dass ein Fremdschlüssel eindeutig der passenden Relation zugeordnet werden kann.

Aufgabe 4 (Fehlersuche) [0.5 Punkte]

Gegeben ist das EER-Diagramm aus Abbildung 2.

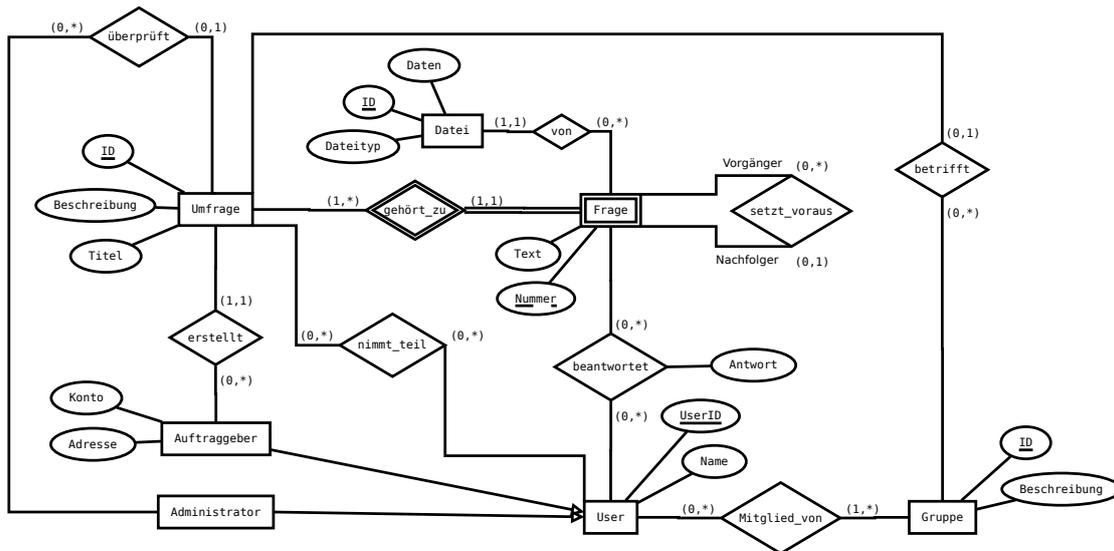


Abbildung 2: EER-Diagramm zu Aufgabe 4

Dazu wurde das folgende (fehlerhafte!) Relationenschema erstellt:

```

User(UserID, Name)
Auftraggeber(User.UserID, Konto, Adresse)
Administrator(User.UserID)
Umfrage(ID, Beschreibung, Titel, User.UserID, Gruppe.ID)
Frage(Umfrage.ID, Nummer, Text, Datei.ID)
Datei(ID, Dateityp, Daten)
Gruppe(ID, Beschreibung)
überprüft(Umfrage.ID, Administrator.UserID)
nimmt_teil(Umfrage.ID, UserID)
setzt_voraus(Frage.ID, Frage.Nummer, Frage.ID, Frage.Nummer)
beantwortet(User.UserID, Frage.ID, Frage.Nummer, Antwort)
Mitglied_von(User.UserID, Gruppe.ID)

```

Finden Sie mindestens 4 Fehler, die sich bei der Überführung des EER-Diagramms ins Relationenmodell eingeschlichen haben, und korrigieren Sie das Relationenschema entsprechend. Wie lautet das korrekte Relationenschema? Beachten Sie, dass NULL-Werte nicht erlaubt sind.

Relationale Algebra - Relationenkalkül

Um Ihnen die Erstellung Ihrer Abgabe zu den folgenden Aufgaben zu erleichtern, haben wir unter <http://dbai.tuwien.ac.at/education/dm/resources/symbols.html> eine Liste mit den wichtigsten Symbolen der relationalen Algebra zusammengestellt. Sie

können diese per copy/paste in Ihr Word/LibreOffice/OpenOffice/... Dokument einfügen. Zusätzlich sind die entsprechenden L^AT_EX Befehle vermerkt.

Aufgabe 5 (Primitive Operatoren) [0.5 Punkte]

Drücken Sie die Operationen Join (\bowtie), Durchschnitt (\cap), Semi-Join (\ltimes) und Anti Semi-Join (\ltimes) mit Hilfe der primitiven Operatoren (σ , Π , \times , ρ , $-$, \cup) aus. Nehmen Sie an, dass zwei Relationen S und T gegeben sind. Hierbei umfasst das Schema von S $g + s$ viele Attribute ($S_1, \dots, S_s, G_1, \dots, G_g$) und das Schema von T $g + t$ viele Attribute ($T_1, \dots, T_t, G_1, \dots, G_g$).

Beim Mengendurchschnitt (\cap) ist zu beachten, dass die Schemata der Relationen übereinstimmen müssen. Somit ist $S \cap T$ kein gültiger Ausdruck in relationaler Algebra. Deshalb betrachten wir hier $S' \cap T'$ mit $S' := \Pi_{G_1, \dots, G_g}(S)$ und $T' := \Pi_{G_1, \dots, G_g}(T)$.

Beschreiben Sie weiters einen neuen, zusammengesetzten relationalen Operator in natürlicher Sprache und veranschaulichen Sie dessen Anwendung an einem Beispiel. Drücken Sie diesen Operator anschließend mit Hilfe der primitiven Operatoren aus. Der neue Operator soll aus zumindest drei primitiven Operatoren zusammengesetzt werden.

Aufgabe 6 (Äquivalenzen) [0.5 Punkte]

Gegeben sind folgende relationale Ausdrücke über $R(\underline{ABC})$, $S(\underline{DE})$, $T(\underline{BD})$ und $U(\underline{AB})$. Überprüfen Sie, ob die folgenden Ausdrücke in relationaler Algebra äquivalent sind. Falls ja, so erklären Sie warum. Falls nein, geben Sie eine Begründung an.

- (a) $r_1 : R \bowtie S$ und $r_2 : S \times R$
- (b) $r_3 : \Pi_B(R) \bowtie \Pi_B(T)$ und $r_4 : \Pi_B(R \bowtie T)$
- (c) $r_5 : \Pi_B(R) \bowtie \Pi_B(U)$ und $r_6 : \Pi_B(R \bowtie U)$
- (d) $r_7 : \sigma_{A \neq 3 \vee B > 5}(R \bowtie U)$ und $r_8 : \sigma_{A \neq 3}(R \bowtie U) \cup \sigma_{B > 5}(R \bowtie U)$
- (e) $r_9 : \sigma_{A \neq 3 \wedge B < 5}(R)$ und $r_{10} : \sigma_{A \neq 3}(R) \cup \sigma_{B < 5}(R)$
- (f) $r_{11} : \Pi_D(S) - \Pi_D(T)$ und $r_{12} : \Pi_D(S - T)$

Aufgabe 7 (Größenabschätzung) [0.5 Punkte]

Gegeben sind die Relationen $R(\underline{ABCD})$ mit x Tupeln, $S(\underline{BDE})$ mit y Tupeln und $T(\underline{ACD})$ mit z Tupeln. Weiters sind folgende Ausdrücke in relationaler Algebra gegeben:

$$\begin{array}{ll}
 r_1 : \Pi_B(\rho_{B \leftarrow C}(T)) \cup \Pi_B(R \bowtie S) & r_2 : \Pi_A(S \bowtie T) \bowtie \Pi_A(R) \\
 r_3 : \Pi_A(R \bowtie T) \cup \Pi_A(S \bowtie T) & r_4 : \Pi_A(\rho_{A \leftarrow B}(S) \bowtie T) \cup \Pi_A(R \bowtie T)
 \end{array}$$

Geben Sie die minimale bzw. maximale Größe (= Anzahl der Tupel) der durch die folgenden Annahmen entstehenden Relationen an und begründen Sie Ihre Antwort.

- (a) Betrachten Sie r_1 unter der Annahme $x = 9, y = 12, z = 4$.
- (b) Betrachten Sie r_2 unter der Annahme $x = 9, y = 12, z = 4$.

In den folgenden Unteraufgaben sind keine konkreten Werte für x, y und z gegeben. Berechnen Sie abhängig von den Variablen x, y und z die minimale und maximale Anzahl der Tupel in der Ergebnismenge und begründen Sie Ihre Antwort.

(c) Betrachten Sie r_3 unter der Annahme $x \leq y \leq z$.

(d) Betrachten Sie r_4 unter der Annahme $1 \leq x \leq y \leq z$.

Aufgabe 8 (Formalisieren von Anfragen) [0.5 Punkte]

Gegeben ist eine Personaldatenbank in Form des folgenden Relationenschemas:

M = Mitarbeiter(PNR, Vorname, Nachname, Wohnort)
 A = Abteilung(AID, Name, *Mitarbeiter.PNR*)
 I = arbeitet_in(*Mitarbeiter.PNR*, Abteilung.AID, Zeitraum)

Dabei ist davon auszugehen, dass die in der Relation A gespeicherte PNR die PNR des Abteilungsleiters ist.

Drücken Sie folgende Anfragen in relationaler Algebra aus.

- (a) Gesucht sind die Namen (Vorname und Nachname) der Abteilungsleiter von allen Abteilungen, in denen schon mindestens ein Mitarbeiter gearbeitet hat, der in Wien wohnt.
- (b) Gesucht sind die Namen (Vorname und Nachname) aller Mitarbeiter, die weder in Linz noch in Salzburg wohnen, keine Abteilungsleiter sind und noch nie in der Abteilung „Controlling“ gearbeitet haben.
- (c) Betrachten Sie nun den folgenden Ausdruck in relationaler Algebra. Beschreiben Sie in eigenen Worten, welches Ergebnis dieser Ausdruck liefert.

$$\Pi_{\text{Vorname, Nachname}} \left(M \times \left(\Pi_{\text{PNR, AID}}(I) \div \Pi_{\text{AID}}(A \times \sigma_{\text{Wohnort}='Wien'}(M)) \right) \right)$$

Aufgabe 9 (Relationenkalkül) [0.5 Punkte]

Übersetzen Sie die Anfragen, die Sie in Aufgabe 8 (a) und Aufgabe 8 (b) in relationaler Algebra formuliert haben, in sichere Ausdrücke des Tupel- und Domänenkalküls. Geben Sie bei jeder der Anfragen an, in welchem Kalkül sie formuliert ist.

Gesamter Stoff

Aufgabe 10 (Falsche Merksätze?) [0.5 Punkte]

Ihr Kollege Karl Konfus erklärt einige wichtige Zusammenhänge, die er sich für die Prüfung merken will. Leider hat er manche Dinge falsch verstanden. Finden Sie die Fehler, begründen Sie warum manche Erklärungen richtig oder falsch sind, und stellen Sie die falschen Aussagen richtig.

- (a) Angenommen, es ist eine Relation **A** mit Schema \underline{CD} und eine Relation **B** mit Schema \underline{CE} gegeben. Der Ausdruck $A \cup B$ in der relationalen Algebra liefert alle Tupel, die in mindestens einer der beiden Relationen vorkommen.
- (b) Mit der (min, max)-Notation kann man in ER-Diagrammen Bedingungen an die Kardinalitäten ausdrücken, die in der M:N-Notation nicht ausdrückbar sind.
- (c) Variablen im Relationenkalkül stehen, wie der Name schon sagt, immer für Relationen. Wenn ich ausdrücken will, dass es in der Relation R ein Tupel gibt, das für das Attribut A den Wert 0 hat, kann ich beispielsweise schreiben $\exists R(R.A = 0)$.
- (d) Wenn in einer Relation **R** ein Fremdschlüsselattribut vorkommt, ist dieses Attribut immer Teil des Schlüssels und gehört daher im Relationenschema unterstrichen.
- (e) Angenommen, **B** ist ein schwacher Entitätstyp, der von dem übergeordneten Entitätstyp **A** abhängt. Dann ist bei der Beziehung zwischen **A** und **B** auf der Seite von **B** nur die Kardinalität (1, 1) möglich.