

1. Übungsblatt

3.0 VU Datenmodellierung

24. April 2012

Allgemeines

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

Lösen Sie die Beispiele eigenständig, denn in der Praxis (und bei der Prüfung) 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

04.05. 06:55 Uhr Upload über den COURSEMANAGER

11.05. 12:00 Uhr Feedback im COURSEMANAGER verfügbar

13.05. 06:55 Uhr Reservierung eines Termins für das Abgabegespräch

Abgabegespräch

1. Sie müssen sich über den COURSEMANAGER zu einem Abgabegespräch (=Prüfungsgespräch) anmelden. Bitte machen Sie das rechtzeitig, je später Sie sich anmelden, umso eingeschränkter ist das Terminangebot. Bis zur oben genannten Deadline garantieren wir Ihnen einen Termin.
2. Sie müssen mindestens einen Punkt auf das Übungsblatt bekommen. Wenn Sie weniger als einen Punkt auf das Blatt bekommen, oder kein Blatt abgegeben 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.

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. Um die Lehrveranstaltung positiv abzuschließen, brauchen Sie mindestens 1 Punkt auf das Übungsblatt und 1 Punkt auf das Abgabegespräch.
8. Falls Sie nicht zu Ihrem Gesprächstermin erscheinen, bekommen Sie automatisch 0 Punkte und damit ein negatives Zeugnis.

ER

Aufgabe 1 (ER Diagramm erstellen) *[0.75 Punkte]* Zeichnen Sie folgenden Sachverhalt in einem ER-Diagramm. Verwenden Sie dazu die (min, max)-Notation. Nullwerte sind nicht erlaubt. Eine Unterstützung bei der Erstellung von ER Diagrammen bietet das Tool dia (<http://live.gnome.org/Dia>). Achtung: im Diagramm Editor ER auswählen!

Mitarbeiter sind eindeutig identifiziert durch ihre Mitarbeiternummer (MNR). Es wird auch noch der Vorname (VNAME), der Nachname (NNAME), die (eindeutige) Sozialversicherungsnummer (SVNR) und das Geschlecht (GESCHLECHT) gespeichert. Optional kann für gekündigte Mitarbeiter ein Kündigungsdatum (DATUM) gespeichert werden. Es gibt zwei unterschiedliche Typen von Mitarbeitern: Vertreter und Kundenbetreuer. Bei Vertretern ist bekannt, welchen Dienstwagen (DIENSTWAGEN) sie benutzen. Bei Kundenbetreuern sind die Telefonnummer (TELEFON) und E-Mail-Adresse (EMAIL) bekannt. Alle Mitarbeiter können mehrere andere Mitarbeiter als Vorgesetzte haben.

Die Produkte der Versicherung sind eindeutig identifiziert durch die Produkt-ID (PID). Zusätzlich werden die Bezeichnung (BEZEICHNUNG), der Preis (PREIS) und die maximal gedeckte Schadenshöhe (MAXSCHADEN) aller Produkte in der Datenbank abgelegt. Ein Produkt gehört jeweils zu einer Kategorie, wobei einer Kategorie beliebig viele Produkte zugeordnet werden können. Bei einer Kategorie sind folgende Eigenschaften bekannt: Eine global einzigartige Kategorie-ID (KATID) sowie die Bezeichnung der Kategorie

(BEZEICHNUNG). Ein Vertreter kann nun beliebig viele Produkte vertreiben und erhält dazu jeweils eine Vorgabe (VORGABE), die besagt, welche Menge er von einem bestimmten Produkt verkaufen sollte.

Kunden sind eindeutig identifiziert durch ihre Kundennummer (KNR). Es werden bei den Kunden der Vorname (VNAME), der Nachname (NNAME) und das Geburtsdatum (GEBDAT) hinterlegt. Kunden können nun im Zuge von Aufträgen Produkte von Vertretern erwerben: Ein Auftrag ist identifiziert durch den jeweiligen Kunden und das gewählte Produkt, sowie das Kaufdatum (DATUM). Zusätzlich wird noch gespeichert, in welcher Höhe der Kunde einen Rabatt (RABATT) auf den Auftrag bekommt, sowie der zuständige Vertreter.

Ein Kunde kann mehrere Bankverbindungen haben, die eindeutig durch die Bankleitzahl (BLZ) und die Kontonummer (KNR) identifiziert sind. Bei Bankverbindungen wird zusätzlich der Name der Bank (BANK) abgespeichert und es ist zu beachten, dass eine Bankverbindung immer zu exakt einem Kunden gehört.

Ein Kundenbetreuer berät jeweils eine beliebige Anzahl von Kunden im Rahmen einer ebenfalls beliebigen Anzahl an Themen. Bei Themen sind jeweils der eindeutige Titel (TITEL) sowie eine Beschreibung (BESCHREIBUNG) bekannt.

Bei auftretenden Schadensfällen können Kunden eine Schadensmeldung einreichen, die durch eine eindeutige ID (ID) identifiziert wird. Weiters werden eine Detailinformation (INFO), das Datum (DATUM) und der geschätzte Gesamtbetrag des Schadens (BETRAG) zu jeder Schadensmeldung gespeichert, sowie der Kunde, den die Meldung betrifft.

Zur Abwicklung eines Schadensfalles wird der Bearbeitungsstatus in der Datenbank abgespeichert. Eine Bearbeitung ist hierbei eindeutig identifiziert durch die Schadensmeldung, der sie zugeordnet ist und als Attribute einer Bearbeitung werden das Datum (DATUM), eine Anmerkung (ANMERKUNG) und der tatsächlich anerkannte Schadensbetrag (BETRAG) gespeichert. Für die Nachverfolgung der Bearbeitung muss auch die verwendete Bankverbindung und der für die Bearbeitung zuständige Mitarbeiter abrufbar sein.

Lösung: Siehe Abbildung 1.

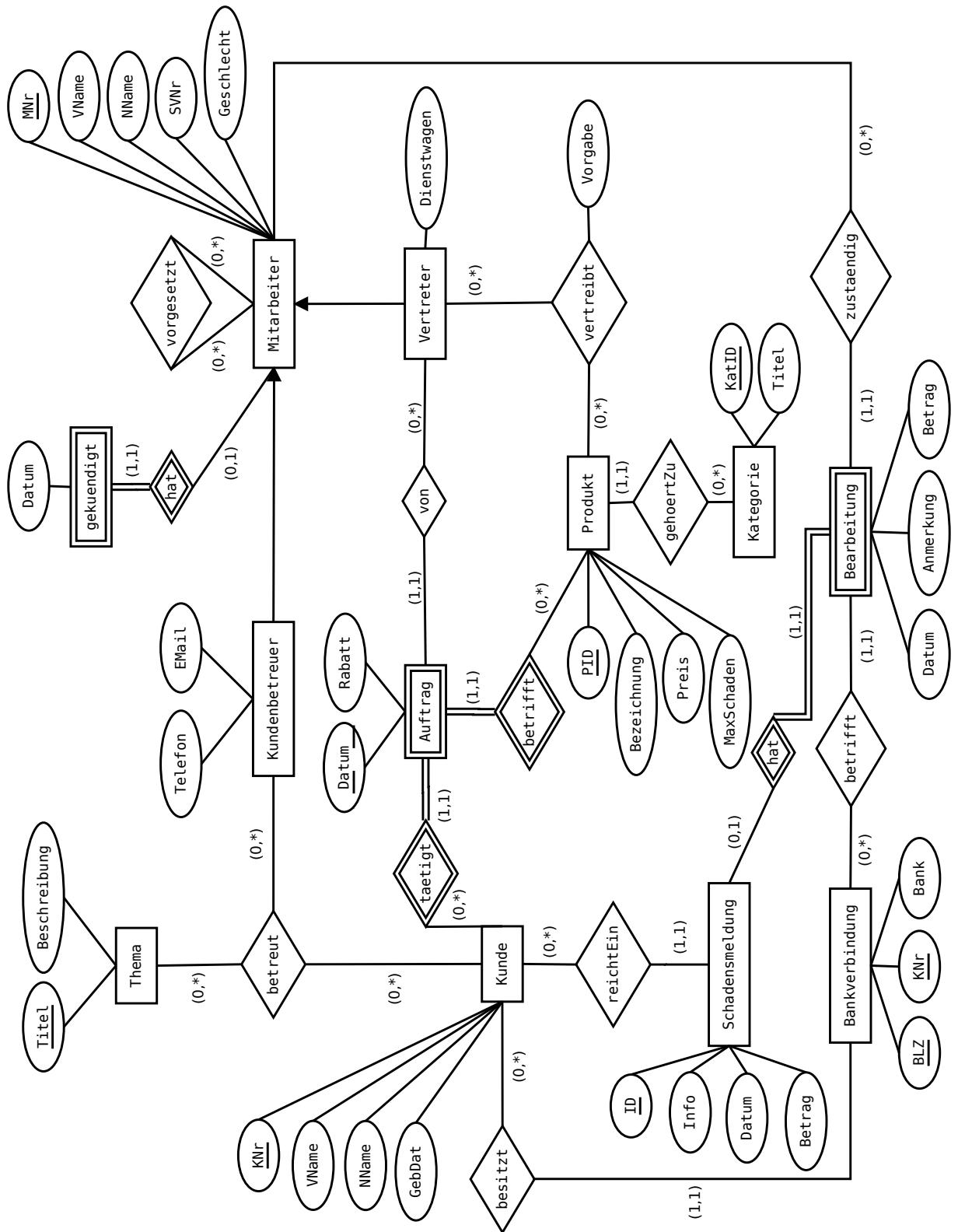


Abbildung 1: Lösung von Aufgabe 1

Aufgabe 2 (Überführung ins Relationenschema) [0.5 Punkte] Überführen Sie folgendes ER-Diagramm (Abbildung 2) ins Relationenschema. Nullwerte sind nicht erlaubt. Beachten Sie hierbei, dass Sie möglichst wenige Relationen verwenden. Unterstreichen Sie sämtliche Primärschlüssel, schreiben Sie die Fremdschlüssel kursiv und stellen Sie sicher, dass ein Fremdschlüssel eindeutig der passenden Relation zugeordnet werden kann.

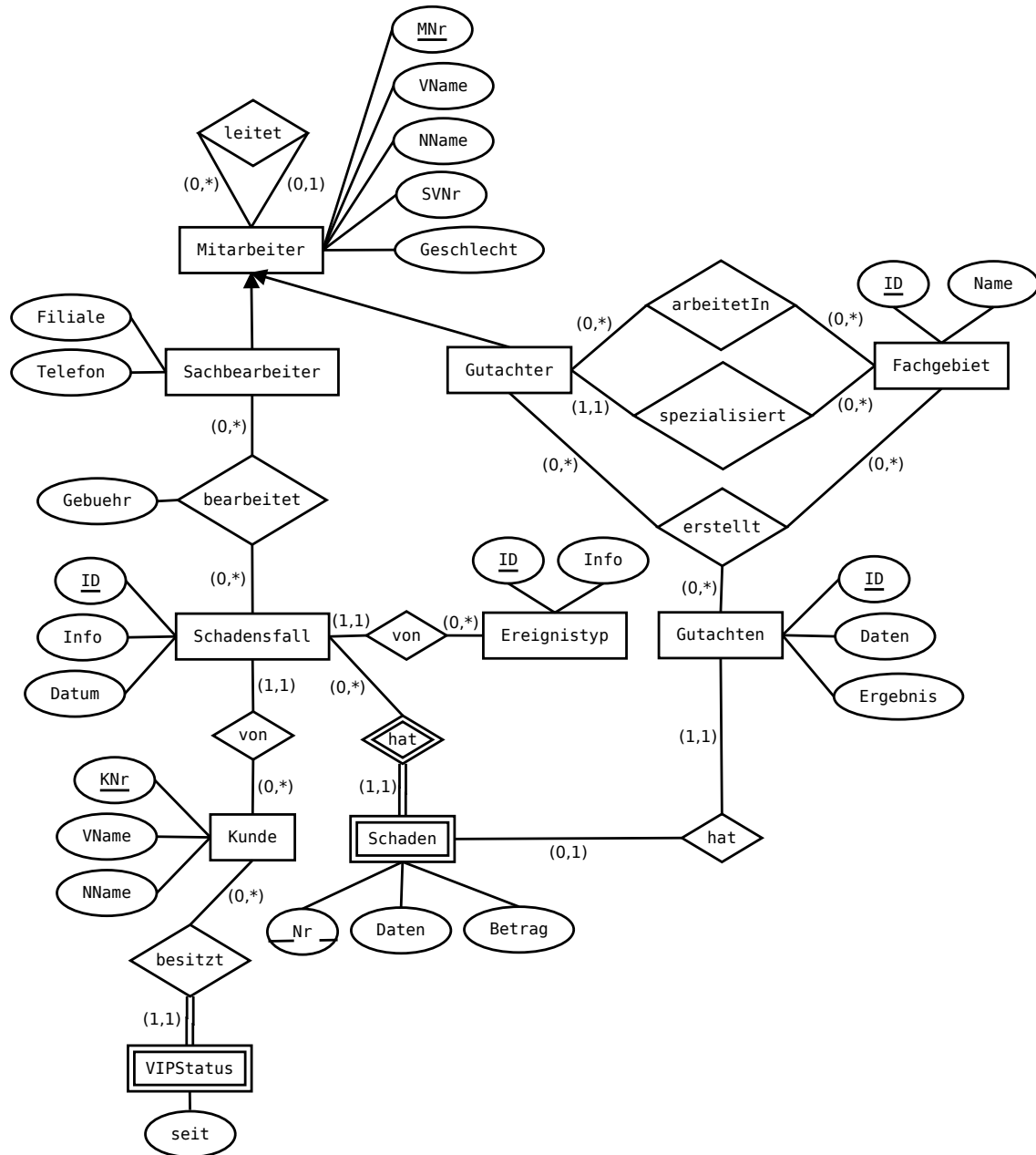


Abbildung 2: ER Diagramm zu Aufgabe 2

Lösung:

```

Fachgebiet(id, name)
Ereignistyp(id, info)
Kunde(knr, vname, nname)
Mitarbeiter(mnr, vname, nname, svnr, geschlecht)
Gutachter(Mitarbeiter.mnr, Fachgebiet.id)
Sachbearbeiter(Mitarbeiter.mnr, filiale, telefon)
Schadensfall(id, info, datum, Ereignistyp.id, Kunde.knr)
VIPStatus(Kunde.knr, seit)
Schaden(Schadensfall.id, nr, daten, betrag)
Gutachten(id, daten, ergebnis, Schaden.id, Schaden.nr)
bearbeitet(Sachbearbeiter.mnr, Schadensfall.id, gebuehr)
erstellt(Gutachten.id, Gutachter.mnr, Fachgebiet.id)
leitet(Mitarbeiter.mnr, Mitarbeiter.mnr)
arbeitetIn(Gutachter.mnr, Fachgebiet.id)

```

Aufgabe 3 (Fehlersuche) [0.5 Punkte] Gegeben ist das folgende ER-Diagramm (Abbildung 3).

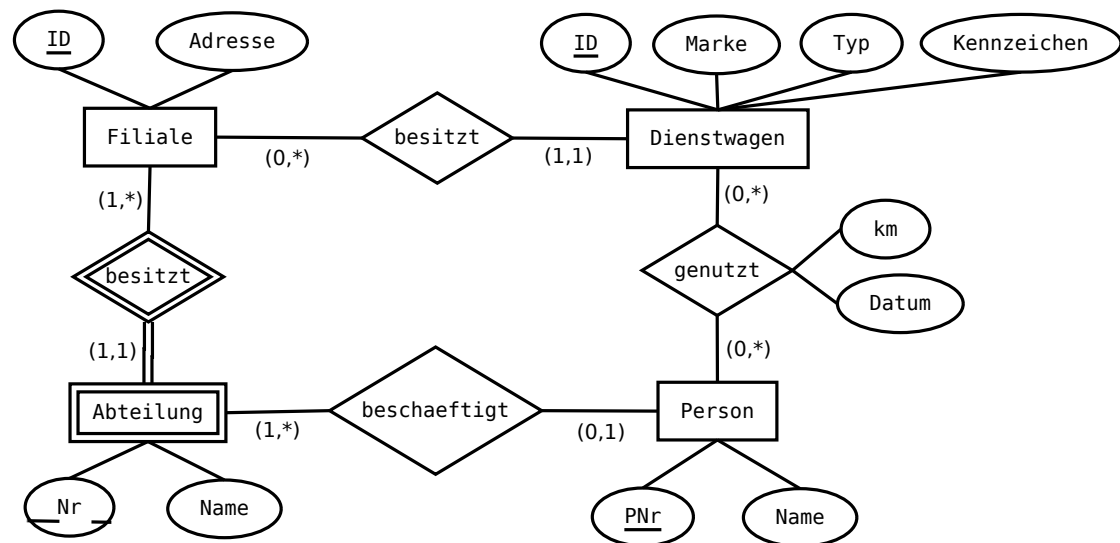


Abbildung 3: ER Diagramm zu Aufgabe 3

Personen haben eine eindeutige Personalnummer (PNR), einen Namen und können bei einer Abteilung angestellt sein.

Von Filialen ist eine eindeutige Identifikation (ID) und ihre Adresse (ADRESSE) bekannt. Filialen bestehen aus mindestens einer Abteilung und besitzen eine beliebige Anzahl an Dienstwagen. Eine Abteilung hat eine pro Filiale ein-

deutige Nummer (NR), sowie einen eindeutigen Namen innerhalb der Filiale (NAME).

Von einem Dienstwagen sollen seine eindeutige Kennung (ID), dessen Marke (MARKE) und Typ (TYP) und das Kennzeichen (KENNZEICHEN) in der Datenbank gespeichert werden können und es muss erkennbar sein, zu welcher Filiale der Wagen gehört.

Dienstwagen können von beliebigen Mitarbeitern für dienstliche Fahrten genutzt werden. Um die Nutzung der Fahrzeuge überwachen zu können, wird gespeichert, welcher Mitarbeiter welchen Wagen verwendet hat. Zusätzlich muss es möglich sein, die insgesamt gefahrenen Kilometer (KM) eines Mitarbeiters mit einem bestimmten Wagen und das jeweilige Datum (DATUM) seiner letzten Dienstreise mit diesem Wagen aus der Datenbank abrufen zu können.

Dazu wurde das folgende (fehlerhafte !) Relationenschema erstellt:

```
Person(PNR, NAME)
Filiale(ID, ADRESSE)
Abteilung(Filiale.ID, NR, NAME)
Dienstwagen(ID, MARKE, TYP, KENNZEICHEN)
besitzt(Filiale.ID, Dienstwagen.ID)
beschaeftigt(Person.PNR, Filiale.ID)
genutzt(Person.PNR, Dienstwagen.ID, KM, DATUM)
```

Finden Sie mindestens 4 Fehler, die sich bei der Überführung des ER-Diagramms ins Relationenmodell eingeschlichen haben, und korrigieren Sie das Relationenschema entsprechend. Wie lautet das korrekte Relationenmodell?

Lösung:

Folgende Fehler wurden korrigiert:

- Der Entitätstyp **Abteilung** wurde nicht korrekt als schwache Entität abgebildet, da das von der Filiale zu übernehmende Schlüsselattribut ID plötzlich nicht mehr Teil des Schlüssels ist.
- Sowohl die Nummer der Abteilung als auch der Name einer Abteilung sind innerhalb einer Filiale eindeutig. Daher muss ein Schlüsselkandidat gewählt werden. Beide Attribute zusammen verletzen die Bedingung, dass ein Schlüssel minimal sein muss.
- Die Beziehung **besitzt** ist einerseits unnötig, da bei Kardinalität (1,1) ohnehin keine NULL-Werte auftreten dürfen. Ganz grundsätzlich ist diese Beziehung nicht korrekt modelliert worden, da es hier möglich ist, dass es nicht für jeden Wagen eine zugehörige Filiale gibt. Zusätzlich wurde auch die Kardinalität in die falsche

Richtung interpretiert, da die aktuelle Umsetzung bedeuten würde, dass es für einen Dienstwagen mehrere Filialen als Besitzer geben kann. Es kann aber pro Filiale höchstens einen Dienstwagen geben. Deshalb wird die ID der Filiale in der Beziehung *Dienstwagen* angeführt.

- Die Beziehung *genutzt* enthält die Attribute KM und DATUM als Schlüsselattribute. Da nun alle Attribute Teil des Schlüssels sind, wurde ein nicht erwünschtes Szenario modelliert. Beispielsweise könnte man mehrere letzte Fahrten mit unterschiedlichen Gesamtkilometerständen für ein und denselben Mitarbeiter speichern.
- In der Beziehung *beschaeftigt* fehlt das Attribut NR der Beziehung *Abteilung*, um die Abteilung eindeutig identifizieren zu können.

Korrigiertes Relationenmodell:

```

Person(PNR, NAME)
Filiale(ID, ADRESSE)
Abteilung(Filiale.ID, NR, NAME)
Dienstwagen(ID, MARKE, TYP, KENNZEICHEN, Filiale.ID)
beschaeftigt(Person.PNR, Filiale.ID, Abteilung.NR)
genutzt(Person.PNR, Dienstwagen.ID, KM, DATUM)

```

Aufgabe 4 (Ternäre Beziehungen eliminieren) [0.5 Punkte] Für folgenden Sachverhalt wurde die folgende ternäre Beziehung (Abbildung 4a) modelliert:

“Produkte können von Kunden über Lieferanten bestellt werden. Eine Bestellung ist einem Kunden zugeordnet und umfasst mindestens ein Produkt und mindestens einen Lieferanten.”

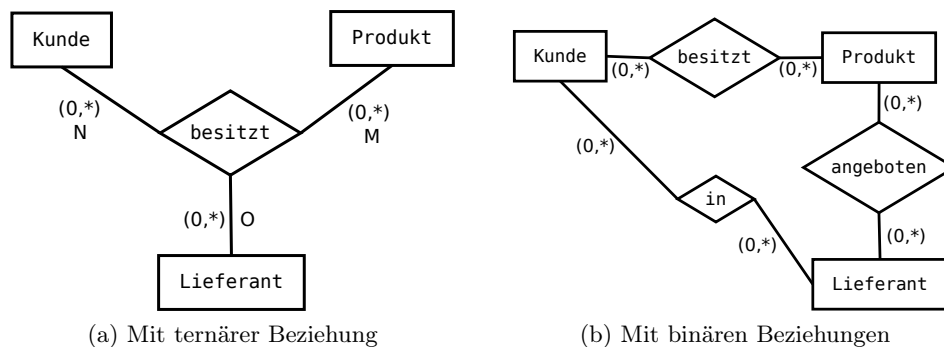


Abbildung 4: ER Diagramme zu Aufgabe 4

Ist in diesem Fall auch eine Modellierung mittels binärer Beziehungen, wie in Abbildung 4b gezeigt, möglich? Begründen Sie Ihre Antwort!

Lösung:

Nein, die gegebene Konvertierung ist nicht korrekt, da die Information verloren geht, über welche Lieferanten der Kunde welche einzelnen Produkte bestellt hat.

Relationale Algebra - Relationenkalkül

Um Ihnen die Erstellung Ihrer Abgabe zu den folgenden Beispielen 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/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) und Semi-Join (\ltimes) mit Hilfe der primitiven Operatoren (σ , Π , \times , ρ , $-$, \cup) aus.

Lösung:

Es wird angenommen, 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 ($T_1, \dots, T_t, G_1, \dots, G_g$).

(a) $S \bowtie T \equiv \Pi_{S.G_1, \dots, S.G_g, S_1, \dots, S_s, T_1, \dots, T_t} (\sigma_{S.G_1=T.G_1 \wedge \dots \wedge S.G_g=T.G_g} (S \times T))$

(b) $S \cap T \equiv S - (S - T)$

(c) $S \ltimes T \equiv \Pi_{S.G_1, \dots, S.G_g, S_1, \dots, S_s} (\sigma_{S.G_1=T.G_1 \wedge \dots \wedge S.G_g=T.G_g} (S \times T))$

Aufgabe 6 (Äquivalenzen) [0.5 Punkte] Gegeben sind folgende Relationale Ausdrücke über $R(\underline{AB})$, $S(\underline{BAC})$ und $T(\underline{DBC})$. 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) $r_1 : \sigma_{B \neq A \wedge B > 10} (\sigma_{B \neq A} (R \cup \Pi_{AB}(S)))$ und $r_2 : \sigma_{B > 10} (\sigma_{B \neq A} (\Pi_{AB}(S) \cup (R \cap R)))$

(b) $r_3 : \Pi_A(R \bowtie T)$ und $r_4 : \Pi_A(R) \bowtie T$

(c) $r_5 : \sigma_{B=13}(T) \cap \sigma_{C=31}(T)$ und $r_6 : \sigma_{B=13 \wedge C=31}(T)$

(d) $r_7 : S \ltimes T$ und $r_8 : T \ltimes S$

(e) $r_9 : \Pi_B(\sigma_{B > 1 \vee C \neq 3}(T))$ und $r_{10} : \Pi_B(\sigma_{B > 1 \vee C \neq 3}(\Pi_{BC}(T)))$

(f) $r_{11} : \Pi_{AB}(S) \bowtie \Pi_{BD}(T)$ und $r_{12} : \Pi_{AB}(S) \times \Pi_{BD}(T)$

Lösung:

(a) **Ja:** Die Konjunktion in der Selektionsbedingung kann aufgebrochen werden. Weiters gilt $R \cap R = R$ und $R \cup S = S \cup R$.

(b) **Nein:** Hier ist zu beachten, dass der Join in r_4 zum Kreuzprodukt degeneriert, was in r_3 nicht der Fall ist. Weiters sind die Schemata unterschiedlich.

- (c) **Ja:** Wenn in Menge A die Bedingung a und in Menge B die Bedingung b gilt, dann erfüllt die Schnittmenge beide Bedingungen.
- (d) **Ja:** Dies ist die selbe Operation in einer anderen Schreibweise.
- (e) **Ja:** Die Projektion auf BC in r_{10} ist unproblematisch, weil alle Attribute, die später in den Selektionsbedingungen betrachtet werden, erhalten bleiben.
- (f) **Nein:** In r_{11} wird hier ein Join über dem Attribut B durchgeführt, wohingegen in r_{12} ein Kreuzprodukt gebildet wird.

Aufgabe 7 (Größenabschätzung) [0.5 Punkte] Gegeben sind die Relationen $R(\underline{A}BC)$ mit 7 Tupeln, $S(\underline{C}D)$ mit 4 Tupeln und $T(\underline{B}DC)$ mit 9 Tupeln.

Geben Sie die minimale bzw. maximale Größe (= Anzahl der Tupel) der durch die folgenden Ausdrücke entstehenden Relationen an:

- (a) $R \bowtie T$ **Lösung:** (min 0/max 9)
- (b) $\Pi_C(S) \times \Pi_B(T)$ **Lösung:** (min 1/max 36)
- (c) $\Pi_{CD}(T) \cap S$ **Lösung:** (min 0/max 4)

Aufgabe 8 (Formalisieren von Anfragen) [0.5 Punkte] Gegeben ist das folgende (stark vereinfachte) Relationenschema:

- $K = \text{Kunde}(\underline{\text{KdnNr}}, \text{VorName}, \text{NachName}, \text{Wohnort}, \text{Alter})$
- $B = \text{borgtAus}(\underline{\text{KdnNr}}, \underline{\text{FilmNr}})$
- $F = \text{Film}(\underline{\text{FilmNr}}, \text{Titel}, \text{Spieldauer})$
- $A = \text{AusgezeichneterFilm}(\underline{\text{FilmNr}}, \underline{\text{Auszeichnung}}, \underline{\text{Jahr}})$

Drücken Sie folgende Anfragen in Relationaler Algebra aus.

- (a) Gesucht sind jene Kunden, die “Resident Evil” ausgeborgt haben, obwohl sie jünger als 16 sind.
- (b) Wir suchen jene erfolglosen Filme, die keine Auszeichnung erhalten haben und von keinem Kunden aus Wien ausgeborgt wurden.

Betrachten Sie nun den folgenden Ausdruck in relationaler Algebra. Beschreiben Sie in eigenen Worten, welches Ergebnis dieser Ausdruck liefert.

(c) $\Pi_{\text{VorName}, \text{NachName}} \left[\left(K \bowtie (B \div \Pi_{\text{FilmNr}}(A)) \right) - \sigma_{\text{Alter} \geq 20}(K) \right]$

Lösung:

- (a) $\sigma_{\text{Alter} < 16}(K) \bowtie (B \bowtie \sigma_{\text{Titel} = \text{'Resident Evil'}}(F))$
- (b) $F \bowtie \left[\left(\Pi_{\text{FilmNr}}(F) - \Pi_{\text{FilmNr}}(A) \right) - \Pi_{\text{FilmNr}} \left(B \bowtie \sigma_{\text{Wohnort} = \text{'Wien'}}(K) \right) \right]$
 oder: $F \bowtie \left[\Pi_{\text{FilmNr}}(F) - \left(\Pi_{\text{FilmNr}}(A) \cup \Pi_{\text{FilmNr}} \left(B \bowtie \sigma_{\text{Wohnort} = \text{'Wien'}}(K) \right) \right) \right]$
- (c) Es werden die Namen jener Filmliebhaber gesucht, die bereits alle Filme mit einer Auszeichnung ausgeborgt haben, aber jünger als 20 Jahre alt sind.

Aufgabe 9 (Relationenkalkül) [0.75 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 Domainkalküls. Geben Sie bei jeder der Anfragen an, in welchem Kalkül sie formuliert ist.

Lösung:

- (a) Tupelkalkül: $\left\{ k \mid k \in K \wedge k.\text{Alter} < 16 \wedge \exists b \in B (k.\text{KdnNr} = b.\text{KdnNr} \wedge \exists f \in F (b.\text{FilmNr} = f.\text{FilmNr} \wedge f.\text{Titel} = \text{'Resident Evil'})) \right\}$

Domänenkalkül: $\left\{ [k, v, n, w, a] \mid [k, v, n, w, a] \in K \wedge a < 16 \wedge \exists f ([k, f] \in B \wedge \exists t, s ([f, t, s] \in F \wedge t = \text{'Resident Evil'})) \right\}$

- (b) Tupelkalkül: $\left\{ f \mid f \in F \wedge \neg \exists a \in A (f.\text{FilmNr} = a.\text{FilmNr}) \wedge \neg \exists b \in B (f.\text{FilmNr} = b.\text{FilmNr} \wedge \exists k \in K (b.\text{KdnNr} = k.\text{KdnNr} \wedge k.\text{Ort} = \text{'Wien'})) \right\}$
 oder z.B.: $\left\{ f \mid f \in F \wedge \forall a \in A (f.\text{FilmNr} \neq a.\text{FilmNr}) \wedge \forall b \in B (f.\text{FilmNr} \neq b.\text{FilmNr} \vee \forall k \in K (b.\text{KdnNr} \neq k.\text{KdnNr} \vee k.\text{Ort} \neq \text{'Wien'})) \right\}$

Domänenkalkül: $\left\{ [f, t, s] \mid [f, t, s] \in F \wedge \neg \exists g, j ([f, g, j] \in A) \wedge \forall k ([k, f] \in B \Rightarrow (\forall v, n, w, a ([k, v, n, w, a] \in K \Rightarrow (w \neq \text{'Wien'}))) \right\}$

oder z.B.: $\left\{ [f, t, s] \mid [f, t, s] \in F \wedge \forall g, j (\neg [f, g, j] \in A) \wedge \neg \exists k ([k, f] \in B \wedge (\exists v, n, w, a ([k, v, n, w, a] \in K \wedge (w = \text{'Wien'}))) \right\}$