

Institut für Informatik Datenbanken und Informationssysteme

29.01.2016

Prof. Dr. Stefan Conrad

Probeklausur

Datenbanksysteme

Hinweise zur Probeklausur Die Probeklausur wird in der Vorlesung am 12.02.2016 besprochen. Die in der Probeklausur enthaltenen Aufgaben sind lediglich Beispiel für mögliche Klausuraufgaben. Dies bedeutet insbesondere, dass sowohl in der Klausur Aufgabentypen vorkommen können, die in der Probeklausur nicht enthalten sind, als auch umgekehrt nicht alle Aufgabentypen aus der Probeklausur in der Klausur vorkommen.

Die Probleklausur wird nicht korrigiert, sie muss deshalb auch nicht abgegeben werden.

Zur Selbstkontrolle: Die Probeklausur entspricht in ihrem Umfang grob einer regulären Klausur, ist also auf etwa 90 Minuten ausgelegt.

Hinweise zu diesem Lösungsvorschlag Diese Lösung ist nur eine beispielhafte Lösung. Es wird weder der Anspruch erhoben, dass die angegebenen Lösungen die einzigen noch die besten Lösungen sind. Weiterhin garantieren wir nicht, dass die hier vorgestellten Lösungen fehlerfrei sind. Ein entsprechender Anspruch auf entsprechende Korrektur in einer Prüfung gilt daher nicht.

Aufgabe 1 ER-Modellierung

(10 Punkte)

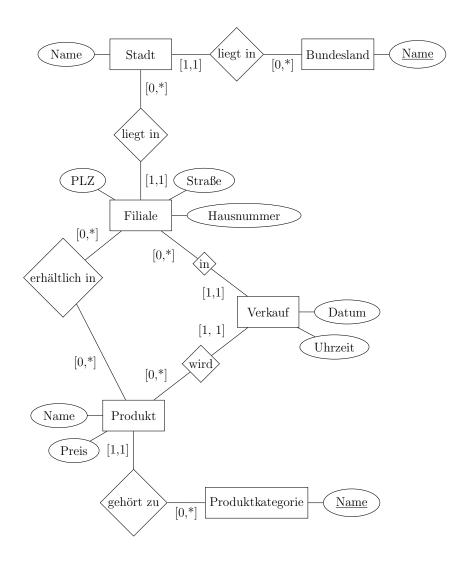
Entwerfen Sie ein ER-Modell zur Verwaltung der Supermarktkette "Centy". Dabei sollen die folgenden Sachverhalte modelliert werden:

- Die Datenbank soll nach Bundesländern und Städten sortiert sein. Dabei liegt eine Stadt in einem Bundesland.
- Filialen liegen in Städten. Außerdem soll zu jeder Filiale die Adresse, bestehend aus Straße, Hausnummer und PLZ gespeichert werden.
- Es wird mit Produkten aus verschiedenen Produktkategorien gehandelt. So ist beispielsweise "Gouda mittelalt" ein Lebensmittel und "Taschentücher extra wolkig" sind eine Haushaltsware. Die Zuordnung von Produkten zu Kategorien soll in der Datenbank möglich sein, die Anzahl der Kategorien ist nicht fix.
- Weiterhin soll zu einem Produkt der Produktname und der Preis des Produktes gespeichert werden.
- Es sind nicht alle Produkte in allen Filialen erhältlich. Deshalb soll die Verfügbarkeit von Produkten in der Datenbank modelliert sein.

• Schließlich soll erfasst werden, welche Produkte an welchem Tag zu welcher Uhrzeit in welcher Filiale verkauft wurden.

Markieren Sie natürliche Schlüssel, falls diese vorhanden sind, durch unterstreichen und geben Sie die Kardinalitäten in Intervallschreibweise an.

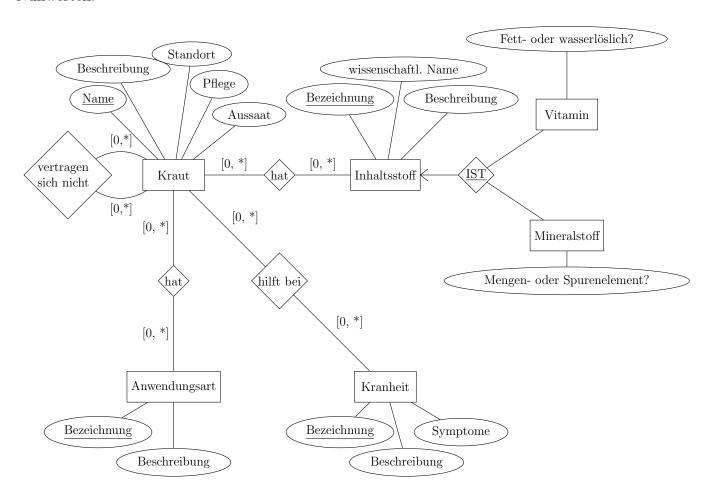
Lösung



Aufgabe 2 Abbildung vom ER-Modell auf das relationale Modell

(10 Punkte)

Gegeben sei das folgende ER-Diagramm, das als Grundlage für eine Kräuter-App dienen soll. Überführen Sie das ER-Diagramm in das relationale Modell. Achten Sie darauf, dass Sie mögliche Verschmelzungen durchführen. Markieren Sie Primärschlüssel durch unterstreichen und Fremdschlüssel durch gestricheltes unterstreichen. Achten Sie bei Ihrer Lösung auf Redundanzfreiheit und Vermeidung von Nullwerten!



Lösung KRAUT (<u>Name</u>, Beschreibung, Standort, Pflege, Aussaat) ANWENDUNGSART (<u>Bezeichnung</u>, Beschreibung) KRANKHEIT (<u>Bezeichnung</u>, Beschreibung, Symptome) INHALTSSTOFF (<u>Bezeichnung</u>, WissenschaftlName, Beschreibung) VITAMIN (<u>Bezeichnung</u>, FettOderWasserlöslich) MINERALSTOFF (<u>Bezeichnung</u>, MengenOderSpurenelement)

KRAUTANWENDUNGSART (Name, Bezeichnung) KRAUTKRANKHEIT (Name, Bezeichnung) KRAUTINHALTSSTOFF (Name, Bezeichnung) VERTRAGENSICHNICHT (Name, Bezeichnung)

Kraut1 und Kraut2 sind Fremdschlüssel auf KRAUT.Name

Die Ist-Beziehung hier ist partiell (es können Instanzen in der Oberklasse existieren, die in keiner Unterklasse vorkommen) und disjunkt (Subklassen sind bzgl. Instanzen zueinander disjunkt). Da wir in der Aufgabenstellung Vermeidung von Nullwerten fordern, kommt nur eine Auflösung der Ist-Beziehungs wie hier angegeben in Frage. Andere bekannte Lösungsmöglichkeiten verursachen Attribute mit Nullwertbelegung.

Aufgabe 3 Anfragesprachen

(5 + 5 + 6 + 4 = 20 Punkte)

Ein Online-Zoofachhandel verwaltet seine Kunden-, Produkt- und Verkaufsdaten mit folgender Datenbank:

KUNDE (Kundennr, Vorname, Nachname, Geburtsdatum, Kunde-seit, Email)

KAUFT (Kundennr, ProduktID, Kaufdatum, Menge, Gesamtpreis)

PRODUKT (ProduktID, Produktname, Preis)

GEEIGNET_FUER (ProduktID, TierartID)

HAUSTIER (Kundennr, TierartID)

TIERART (<u>TierartID</u>, Tierart)

Primärschlüssel sind <u>unterstrichen</u>, Fremdschlüssel sind *kursiv* dargestellt.

- 3.1 Formulieren Sie die folgenden Anfragen in der Relationenalgebra:
 - (a) Welche Produkte (Produktname, Preis) kosten mehr als 100 Euro?

Lösung $\pi_{\text{Produktname. Preis}}(\sigma_{\text{Preis}>100}(\text{PRODUKT}))$

(b) Welche Kunden (Kundennr, Vorname, Nachname) haben eine Ratte als Haustier?

```
Lösung \pi_{\text{Kundennr, Vorname, Nachname}}(\text{KUNDE} \bowtie \text{HAUSTIER} \bowtie (\sigma_{\text{Tierart} = 'Ratte'}(\text{TIERART}))) \pi_{\text{HID,HName}}(\sigma_{\text{PName}='\text{Lasagne'}}(\text{HERSTELLER} \bowtie \text{PRODUKT}))
```

(c) Welche Produkte (Produktname, Preis) wurden noch nie verkauft?

```
Lösung \pi_{\text{Produktname, Preis}}(\text{PRODUKT} \bowtie (\pi_{\text{ProduktID}}(\text{PRODUKT}) - (\pi_{\text{ProduktID}}(\text{KAUFT}))))
```

Achtung: Für die Differenz muss das gleiche Relationenschema vorhanden sein!

(d) Welche Produkte (Produktname, Preis) sind sowohl für Hunde als auch für Katzen geeignet?

Lösung Mittels Division:

```
\pi_{Produktname, Preis}(PRODUKT \bowtie (GEEIGNET\_FUER \div \pi_{TierartID}(\sigma_{Tierart = 'Hund' \lor Tierart = 'Katze'}(TIERART))))
```

Mittels Durchschnitt:

```
\begin{split} &\pi_{\text{Produktname, Preis}}(\text{PRODUKT}\bowtie\\ &(\pi_{\text{ProduktID}}(\text{GEEIGNET\_FUER}\bowtie\sigma_{\text{Tierart}\ =\ '\text{Hund'}}(\text{TIERART}))\cap\\ &\pi_{\text{ProduktID}}(\text{GEEIGNET\_FUER}\bowtie\sigma_{\text{Tierart}\ =\ '\text{Katze'}}(\text{TIERART})))) \end{split}
```

Achtung: Für den Durchschnitt muss das gleiche Relationenschema vorhanden sein! Außerdem ist die Projektion der "zu durchschneidenden" Mengen auf ProduktID notwendig, sonst erhalten wir ein falsches Ergebnis.

- 3.2 Formulieren Sie die folgenden Anfragen mittels Tupelkalkül:
 - (a) Welche Produkte (Produktname, Preis) sind für Meerschweinchen geeignet?

Lösung {p.Produktname, p.Preis | PRODUKT(p)
$$\land$$
 GEEIGNET_FUER(gf) \land TIERART(t) \land p.ProduktID = gf.ProduktID \land gf.TierartID = t.TierartID \land t.Tierart = 'Meerschweinchen'}

(b) Welche Produkte (Produktname, Preis) wurden noch nie verkauft?

Lösung

```
\{p.Produktname, p.Preis \mid PRODUKT(p) \land \neg \exists ka (KAUFT(ka) \land ka.ProduktID) = p.ProduktID)\}
```

(c) Für welche Tierarten wurden 2013 von Kunden, die sich vor 2007 angemeldet haben, schon Produkte gekauft?

Lösung

(d) Für welche Tierarten gibt es Produkte, die auch für Lachtauben geeignet sind?

Lösung

Wiederholung des Schemas:

KUNDE (Kundennr, Vorname, Nachname, Geburtsdatum, Kunde_seit, Email)

KAUFT (Kundennr, ProduktID, Kaufdatum, Menge, Gesamtpreis)

PRODUKT (ProduktID, Produktname, Preis)

GEEIGNET_FUER (ProduktID, TierartID)

HAUSTIER (Kundennr, TierartID)

TIERART (TierartID, Tierart)

- **3.3** Formulieren Sie die folgenden Anfragen in SQL:
 - (a) Für welche Tiere (Tierart) ist das Produkt "Plüschgeier Fred" geeignet?

Lösung

(b) Welche Kunden (Kundennr, Vorname, Nachname) haben in diesem Jahr (2013) noch keine Einkäufe getätigt?

Lösung

```
SELECT Kundennr, Vorname, Nachname
FROM KUNDE
WHERE Kundennr NOT IN (SELECT Kundennr
FROM KAUFT
WHERE Kaufdatum >= '2013-01-01' AND
Kaufdatum < '2014-01-01')
```

(c) Geben Sie die Produkte (ProduktID, Produktname, Preis) mit ihren Gesamtverkaufszahlen (nach Gesamtverkaufszahlen absteigend sortiert) an. Beachten Sie hierbei auch die Menge des verkauften Produkts!

Lösung

```
SELECT p.ProduktID, p.Produktname, p.Preis, SUM(ka.Menge) AS
Gesamtverkaufszahl
FROM KAUFT ka, PRODUKT p
WHERE ka.ProduktID = p.ProduktID
GROUP BY p.ProduktID, p.Produktname, p.Preis
ORDER BY Gesamtverkaufszahl DESC
```

(d) Welche Kunden (Kundennr, Vorname, Nachname) haben mit allen ihren Einkäufen im Jahre 2012 einen Gesamtumsatz von mindestens 500 Euro gemacht? Geben Sie die Kunden und deren Gesamtumsatz (nach Gesamtumsatz absteigend sortiert) an.

Lösung

SELECT ku.Kundennr, ku.Vorname, ku.Nachname, SUM(ka.Gesamtpreis) AS Gesamtumsatz

FROM KAUFT ka, KUNDE ku

WHERE ka.Kundennr = ku.Kundennr AND

ka.Kaufdatum >= '2012-01-01' AND

ka.Kaufdatum <= '2012-12-31'

GROUP BY ku.Kundennr, ku.Vorname, ku.Nachname

HAVING SUM(ka.Gesamtpreis) >= 500

ORDER BY Gesamtumsatz DESC

- **3.4** Formulieren Sie die folgenden Änderungsoperationen in SQL:
 - (a) Ein neuer Kunde Mona Krümel hat sich heute im Onlineshop registriert und die Daten 08.01.1973, mona@kruemel.de hinterlassen, sowie die Haustiere Fisch und Vogel angegeben. Fügen Sie die neuen Daten in die Datenbank ein. Nehmen Sie an, dass die automatisch generierte Kundennr für diese Kundin 20032013128 und die TierartIDs für Fisch und Vogel 4 und 5 lauten.

Lösung

(b) In dieser Woche soll ein Angebot für alle Hundeartikel gelten. Reduzieren Sie die Preise aller Produkte, die für Hunde geeignet sind, um 20~%.

Lösung

(c) Löschen Sie die Produkte, die noch nie verkauft worden sind, aus dem Sortiment.

Lösung

```
DELETE FROM PRODUKT
WHERE ProduktID NOT IN (SELECT ProduktID
FROM KAUFT)
```

Aufgabe 4 Normalformen, Synthese-Algorithmus

$$(2+4+8+6=20 \text{ Punkte})$$

Gegeben sei die Relation S(A,B,C,D,E,F,G,H), für die folgende funktionale Abhängigkeiten gelten:

$$\begin{array}{ccc} D & \rightarrow F \\ AE & \rightarrow CB \\ CG & \rightarrow B \\ C & \rightarrow G \\ E & \rightarrow GA \\ FB & \rightarrow H \end{array}$$

(a) Bestimmen Sie für die Relation S alle Schlüsselkandidaten. Begründen Sie, warum es sich hier um Schlüsselkandidaten handelt.

Lösung Der einzige Schlüsselkandidat ist DE, da

$$\{C\}^+ = \{BCG\}$$

$$\{D\}^+ = \{DF\}$$

$${E}^+ = {ABCEG}$$

$$\{DE\}^+ = \{ABCDEFGH\} = S$$
 und minimal

Alle weiteren erfüllen nicht mehr die Minimalität, da D und E immer benötigt werden.

(b) In welcher Normalform (bis einschließlich BCNF) befindet sich die Relation S? Begründen Sie Ihre Antwort!

Lösung 1. NF: Ja.

- 2. NF: Nein, da es eine partielle Abhängigkeit D \rightarrow F gibt (ein Teil des Schlüssels bestimmt ein Nicht-Prim-Attribut).
- 3. NF: Nein, da es eine transitive Abhängigkeit AE \rightarrow C \rightarrow G gibt oder weil 2. NF shon nicht erfüllt wird.

BCNF: Nein, da nicht in 3. NF

(c) Wenn die Relation S nicht in 3NF ist, überführen Sie diese in die 3. Normalform. Zeigen Sie, dass Ihre Lösung wirklich in 3NF ist.

Lösung Linksreduktion:

 $D \to F$

 $AE \rightarrow CB$ wird zu $E \rightarrow CB$, da $E \rightarrow A$

 $CG \rightarrow B$ wird zu $C \rightarrow B$, da $C \rightarrow G$

 $C \to G$

 $E \to GA$

 $FB \to H$

Rechtsreduktion:

 $D \to F$

 $E \to CB$ wird zu $E \to C$, da $C \to B$

 $C \to B$

 $C \to G$

 $E \to GA$

 $FB \to H$

Zusammenfassen:

 $C \to BG$

 $D \to F$

 $E \to ACG$

 $\mathrm{FB} \to \mathrm{H}$

Damit ergeben sich die Relationen:

 $R1(\underline{C}BG)$, $R2(\underline{D}F)$, $R3(\underline{E}ACG)$, $R4(\underline{F}BH)$ und die Schlüsselrelation $R5(\underline{D}E)$)

(d) Überprüfen Sie, ob die folgenden Dekompositionen für die Relation S verbundtreu und abhängigkeitsbewahrend sind. Begründen Sie Ihre Antwort!

(1) $S_1(CGB)$, $S_2(ACDEFH)$

Lösung Die Zerlegung ist verbundtreu, da BCG\(\)ACDEFH = C und $C^+ = \{BCG\} = S_1$

Die Zerlegung ist nicht abhängigkeitsbewahrend, da die Abhängigkeit FB \to H in keiner Relation vorhanden ist.

(2) $S_1(\text{FBH}), S_2(\text{ABCDEFH})$

Lösung Es sind nicht alle Attribute vorhanden (G fehlt), damit erübrigt sich die Frage nach Verbundtreue und Abhängigkeitsbewahrung.

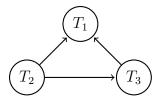
(3) $S_1(DF)$, $S_2(CGB)$, $S_3(EAC)$, $S_4(FBH)$

Lösung Die Zerlegung ist nicht verbundtreu, da S_1 und S_4 kein Attribut mit S_2 und S_3 teilen.

Die Zerlegung ist abhängigkeitsbewahrend, da alle Abhängigkeiten in einer eigenen Relation vorhanden sind.

- (a) Testen Sie die folgenden Schedules auf Ihre Konfliktserialisierbarkeit hin. Zeichnen Sie jeweils den Konfliktgraphen und geben Sie für jede Kante in jedem Graphen jeweils ein Konfliktpaar an.
 - $S_1 := r_1(x), r_2(y), w_1(x), r_2(z), r_3(z), w_2(y), w_3(y), w_1(y)$

Lösung



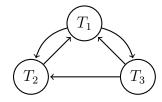
Konfliktpaare:

Konflikt	Paar
$T_2 \rightarrow T_1$	$r_2(y), w_1(y)$
$T_2 \rightarrow T_3$	$r_2(y), w_3(y)$
$T_3 \rightarrow T_1$	$w_3(y), w_1(y)$

Der Schedule ist serialisierbar, da der Konfliktgraph kreisfrei ist.

•
$$S_2 := r_1(z), r_1(x), r_3(z), w_3(z), r_2(x), r_2(z), w_1(x), r_3(y), w_1(z), w_2(y), w_2(z)$$

Lösung



Konfliktpaare:

Konflikt	Paar
$T_1 \rightarrow T_2$	$r_1(z), w_2(z)$
$T_1 \rightarrow T_3$	$r_1(z), w_3(z)$
$T_2 \rightarrow T_1$	$r_2(x), w_1(x)$
$T_3 \rightarrow T_1$	$w_3(z), w_1(z)$
$T_3 \rightarrow T_2$	$w_3(z), r_2(z)$

Der Schedule ist nicht serialisierbar, da der Graph nicht kreisfrei ist: Es existieren die folgenden drei Kreise im Graphen: (T_1, T_2) , (T_1, T_3) , (T_1, T_3, T_2) .

- (b) Gegeben seien die Transaktionen $T_1 := r_1(x), r_1(y), w_1(x), r_1(x)$ und $T_2 := r_2(y), w_2(x), w_2(y)$. Geben Sie für die folgenden Ausdrücke an, ob es sich um Schedules zur Ausführung von T_1 und T_2 handelt. Begründen Sie Ihre Antwort, falls es sich nicht um einen Schedule handelt.
 - (1) $S_1 := r_1(x), r_1(y), r_2(y), w_2(x), w_2(y)$

Lösung Es handelt sich nicht um einen Schedule, da nicht alle Operationen von T_1 und T_2 in S_1 vorkommen. So fehlt bspw. $w_1(x)$.

(2)
$$S_2 := r_1(x), r_2(y), r_1(y), w_1(x), w_2(x), r_1(x), w_2(y)$$

Lösung Es handelt sich bei S_2 um einen Schedule, da alle Operationen beider Transaktionen in der richtigen Reihenfolge bearbeitet werden.

(3)
$$S_3 := r_1(x), r_2(y), r_1(y), w_1(x), w_2(y), r_1(x), w_2(x)$$

Lösung Es handelt sich hierbei nicht um einen Schedule, da die beiden letzten Operationen von T_2 in S_3 vertauscht sind: $w_2(y), w_2(x)$ in S_3 gegen $w_2(x), w_2(y)$ in T_2 .

(c) Geben Sie für den folgenden Schedule, wenn möglich, alle äquivalenten seriellen Schedules an:

$$T_1 := r_1(x), x = 10x, r_1(y), y = y + x, r_1(z), z = z + y, w_1(z), w_1(x)$$

$$T_2 := r_2(y), x = y/2, w_2(x), r_2(z), z = 2x, w_2(y)$$

$$S_1 := r_2(y), x = y/2 \text{ (zu } T_2), w_2(x), r_1(x), x = 10x \text{ (zu } T_1), r_2(z), r_1(y), y = y + x \text{ (zu } T_1), x = 2x \text{ (zu } T_2), r_1(z), z = z + y \text{ (zu } T_1), w_1(z), w_1(x), w_2(y)$$

Lösung Aufgrund der Konfliktpaare $w_2(x), r_1(x)$ und $r_1(y), w_2(y)$ ergibt sich der folgende Konfliktgraph:



Dieser Graph ist nicht kreisfrei, daher ist S_1 nicht serialisierbar und es kann daher keinen äquivalenten seriellen Schedule geben.

(d) Ergänzen Sie in folgenden Transaktionen Read-Lock (rl)-, Write-Lock (wl)- und Unlock (u)- Anweisungen nach dem konservativen Zwei-Phasen-Sperrprotokoll. Achten Sie darauf, dass Sperren möglichst früh wieder freigegeben werden.

$$T_1 := r_1(x), w_1(x), w_1(y)$$

 $T_2 := r_2(x), r_2(y), w_2(x)$

Lösung

$$T_1 := wl_1(x), wl_1(y), r_1(x), w_1(x), u_1(x), w_1(y), u_1(y)$$

$$T_2 := wl_2(x), rl_2(y), r_2(x), r_2(y), u_2(y), w_2(x), u_2(x)$$

(e) Konstruieren Sie aus zwei der folgenden drei Transaktionen einen Schedule, der unter dem Zwei-Phasen-Sperrprotokoll zu einem Deadlock führt. Geben Sie in diesem Schedule die Sperren an und begründen Sie, an welcher Stelle es wieso zu einem Deadlock kommt.

$$T_1 := r_1(x), r_1(y), r_1(x), w_1(z), w_1(y)$$

$$T_2 := r_2(x), r_2(y), w_2(x), r_2(z)$$

$$T_3 := r_3(z), r_3(x), r_3(y), w_3(y), w_3(x)$$

Lösung Wähle die Transaktionen T_1 und T_3 . Das Schedule

$$S := rl_1(x), r_1(x), rl_1(y), r_1(y), r_1(x), rl_3(z), r_3(z), rl_3(x), r_3(x), rl_3(y), r_3(y), wl_3(y), wl_3(y), wl_3(x), wl_3(x), ul_3(x), ul$$

führt zu einer Verklemmung, da T_3 an der Stelle $wl_3(y)$ versucht, ein Write-Lock auf y zu bekommen. Dieses kann aber nicht gesetzt werden, da T_1 schon ein Read-Lock auf y hat. An dieser Stelle kann also nur mit T_1 weiter gemacht werden. Auch T_1 kann aber nur bis zu der Stelle fortgeführt werden, zu der versucht wird, ein Write-Lock auf z für T_1 zu bekommen, obwohl T_3 schon ein Read-Lock auf z hat. An dieser Stelle tritt der Deadlock auf.