

3. Übungszettel (Abgabe 12.05.2008)

Einführung in Datenbanksysteme
Datenbanken für die Bioinformatik

Heinz Schweppe, Jürgen Broß, Manuel Scholz

Übungsaufgaben

1. Aufgabe (Generalisierung)

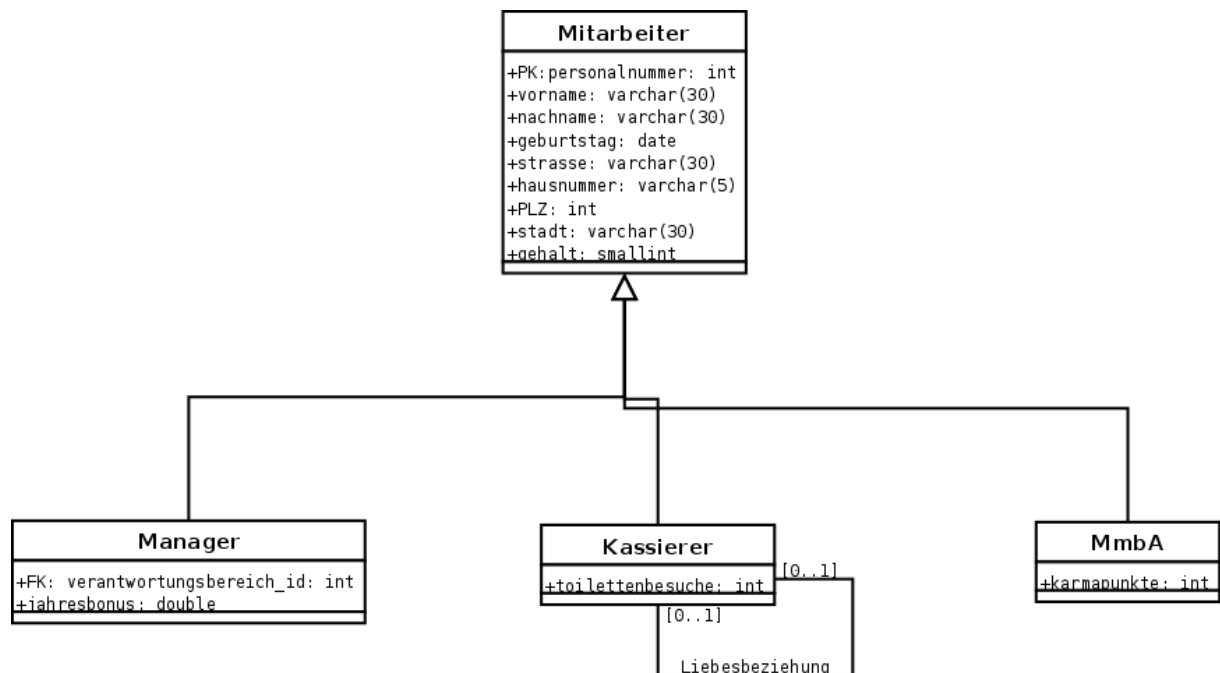
- a) Entwerfen Sie für folgendes Szenario ein ER-Modell (in UML-Notation):

Sie entwickeln gerade eine Datenbank für die Personalabteilung der Supermarktkette „BigL“. Die Firma beschäftigt Manager, Kassierer und „Mitarbeiter mit besonderen Aufgaben (MmbA)“. Die Datenbank soll für alle Mitarbeiter Stammdaten wie etwa Name, Adresse und Gehalt erfassen.

Bei Managern wird zusätzlich zu den Stammdaten noch der Verantwortungsbereich, der Jahresbonus und die ausgehandelte Abfindung erfasst.

Die Anforderungen der Firma „BigL“ verlangen, dass bei Kassierern jeweils noch gespeichert wird wie häufig im Tagesdurchschnitt die Toilette besucht wird und mit welchem anderen Kassierer eine Liebesbeziehung besteht. Ein Kassierer kann maximal eine Liebesbeziehung haben.

MmBa kümmern sich um die Aufrechterhaltung des guten Arbeitsklimas. Dazu beobachten sie ihre Kollegen und machen sich Notizen. Für jede relevante Notiz wird einem *MmBa* ein Karmapunkt in der Personaldatenbank gutgeschrieben.

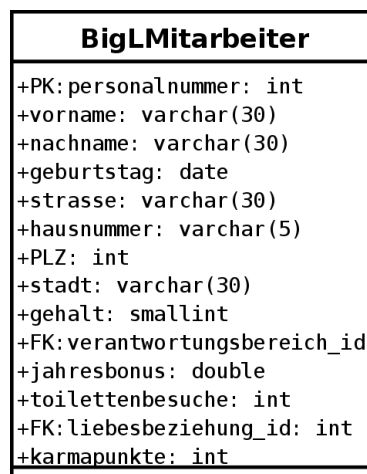


- b) Überführen Sie nun Ihr ER-Modell in ein relationales Schema? Welche 3 Möglichkeiten haben Sie die Generalisierung zu implementieren? Geben Sie für jede Alternative die Vor- und Nachteile an.

Eine große Tabelle

Vorteile: kein Join notwendig

Nachteile: viele Nullwerte, Verlust an Information (z.B. hier: Ein Kassierer kann nur Liebesbeziehung zu einem anderen Kassierer haben)



Separate Tabellen für Obertyp und Spezialisierungen

Vorteile: keine Nullwerte, kein Informationsverlust, flexibel bei Erweiterungen (z.B. neuer Stellentyp: Überwachungsspezialist)

Nachteile: Join notwendig (nicht wirklich ein Nachteil...)

Mitarbeiter
+PK:personalnummer: int
+vorname: varchar(30)
+nachname: varchar(30)
+geburtstag: date
+strasse: varchar(30)
+hausnummer: varchar(5)
+PLZ: int
+stadt: varchar(30)
+gehalt: smallint

Manager
+PK/FK:personalnummer: int
+FK: verantwortungsbereich_id: int
+jahresbonus: double

Kassierer
+PK/FK:personalnummer: int
+toilettenbesuche: int
+FK:liebesbeziehung_id: int

MmbA
+PK/FK:personalnummer: int
+karmapunkte: int

Separate Tabelle für jede Spezialisierung

Vorteile: keine Nullwerte, kein Informationsverlust

Nachteile: Redundanz auf Metadatenebene, Unflexibel (z.B. Hinzufügen einer Stammdatenspalte müsste in jeder Tabelle nachgezogen werden), keine Entitäten speicherbar, die vom Obertyp sind (Spezialisierung muss vollständig sein)

Manager
+PK:personalnummer: int
+vorname: varchar(30)
+nachname: varchar(30)
+geburtstag: date
+strasse: varchar(30)
+hausnummer: varchar(5)
+PLZ: int
+stadt: varchar(30)
+gehalt: smallint
+FK:verantwortungsbereich_id: int
+jahresbonus: double

Kassierer
+PK:personalnummer: int
+vorname: varchar(30)
+nachname: varchar(30)
+geburtstag: date
+strasse: varchar(30)
+hausnummer: varchar(5)
+PLZ: int
+stadt: varchar(30)
+gehalt: smallint
+toilettenbesuche: int
+FK:liebesbeziehung_id: int

MmbA
+PK:personalnummer: int
+vorname: varchar(30)
+nachname: varchar(30)
+geburtstag: date
+strasse: varchar(30)
+hausnummer: varchar(5)
+PLZ: int
+stadt: varchar(30)
+gehalt: smallint
+karmapunkte: int

2. Aufgabe (Funktionale Abhängigkeiten)

Gegeben sei eine Relation R(A,B,C,D):

A	B	C	D
1	1	'music'	3
1	2	'music'	4
2	2	'music'	4

- Welche funktionalen Abhängigkeiten können für das Relationsschema R vorliegen bzw. welche sind ausgeschlossen. "Triviale" Abhängigkeiten der Form $X \rightarrow X$ bzw. $X \rightarrow Y$ mit $Y \rightarrow X$ müssen nicht berücksichtigt werden.

LÖSUNG:

$AB \rightarrow C$, $DB \rightarrow C$, $ABD \rightarrow C$, $A \rightarrow C$, $B \rightarrow C$, $D \rightarrow C$

$B \rightarrow D$

$AB \rightarrow D$ $D \rightarrow B$ **3. Aufgabe (Funktionale Abhängigkeiten)**

Gegeben sei die folgende Tabelle:

a	b	c	d
a1	b1	c1	d1
a1	b1	c2	d2
a2	b1	c1	d3
a2	b1	c3	d4

mit den folgenden Abhängigkeiten:

I. $a \rightarrow b$ II. $c \rightarrow b$ III. $d \rightarrow abc$ IV. $ac \rightarrow d$

a) Nennen Sie Superschlüssel und Schlüsselkandidaten der Relation.

LÖSUNG: d und ac *Superschlüssel sind alle, die d oder ac enthalten.*

b) Welche der folgenden Zeilen können eingefügt werden, ohne Abhängigkeiten zu verletzen? Geben Sie die Nummer der Abhängigkeit an, wenn Sie das Einfügen einer Zeile verletzt und begründen Sie warum.

(1) a5 b6 c7 d8

(2) a2 b2 c1 d8

(3) a3 b1 c4 d3

(4) a1 b1 c2 d5

LÖSUNG:

(1) OK

(2) *geht nicht (1,2,4)*(3) *geht nicht (3)*(4) *geht nicht (4)*

5. Aufgabe (Normalformen)

- a) Geben Sie eine Menge funktionaler Abhängigkeiten an, für die die Relation $R(\underline{a}, b, c, d)$ in erster aber nicht in zweiter Normalform ist.

LÖSUNG:

$a \rightarrow c$ oder d
oder
 $b \rightarrow c$ oder d

- b) Geben Sie eine Menge funktionaler Abhängigkeiten an, für die die Relation $R(\underline{a}, b, c, d)$ in zweiter aber nicht in dritter Normalform ist.

LÖSUNG:

$ab \rightarrow c$
 $c \rightarrow d$

6. Aufgabe (Normalformen)

Gegeben seien folgende Relationen mit den funktionalen Abhängigkeiten:

$R1(A, B, C), \quad F(R1) = \{ \{A, B\} \rightarrow \{C\} \}$
 $R2(A, B, D, E, F), \quad F(R2) = \{ \{E\} \rightarrow \{F\}, \{F\} \rightarrow \{A\}, \{A, B, D\} \rightarrow \{E\} \}$

- a) Bestimmen Sie die Schlüsselkandidaten für $R1$ und $R2$. Bestimmen Sie zu jeder Relation den Schlüssel und begründen Sie Ihre Wahl.

LÖSUNG:

$R1: AB$
 $R2: ABD, EBD, FBD$

- b) In welcher Normalform sind $R1$ bzw. $R2$? Begründen Sie Ihre Antwort.

LÖSUNG:

$R1: BCNF$
 $R2: 3. NF$