TU München, Fakultät für Informatik
Lehrstuhl III: Datenbanksysteme
Prof. Alfons Kemper, Ph.D.

# Übung zur Vorlesung *Grundlagen: Datenbanken* im WS13/14

Henrik Mühe (muehe@in.tum.de)
http://www-db.in.tum.de/teaching/ws1314/dbsys/exercises/

**Blatt Nr. 9**

## Hausaufgabe 1

Gegeben sei eine Relation

$$\mathrm{R} \;:\; \{[\mathrm{A} : \mathrm{integer}, \mathrm{B} : \mathrm{integer}, \mathrm{C} : \mathrm{integer}, \mathrm{D} : \mathrm{integer}, \mathrm{E} : \mathrm{integer}]\},$$

die schon sehr viele Daten enthält (Millionen Tupel). Sie „vermuten“, dass folgendes gilt:

(a) $AB$ ist ein Schlüssel der Relation

(b) $DE \rightarrow B$

Formulieren Sie SQL-Anfragen, die Ihre Vermutungen bestätigen oder widerlegen.

(a) Da jedes Tupel in einer Relation einen eindeutigen Schlüssel besitzt, kann nach der Gruppierung nach $A$ und $B$ anhand der Anzahl der Tupel ermittelt werden, ob hier eine Verletzung der Schlüsseleigenschaft vorliegt. Werden also mindestens zwei Tupel mit den gleichen Werten für $A$ und $B$ als Ergebnis ausgegeben, so bildet $AB$ keinen Schlüssel der Relation, ist das Ergebnis der Anfrage leer, so ist $AB$ ein Schlüssel.

```
select A, B
from R
group by A, B
  having count(*) > 1;
```

(b) In diesem Fall muss nur gelten, dass für alle Tupel, die gleiche Werte in $D$ und $E$ besitzen, auch die Werte für das Attribut $B$ gleich sind. D.h. wenn nach $D$ und $E$ gruppiert wird, muss die Anzahl der verschiedenen Werte für $B$ kleiner oder gleich 1 sein. Es gilt wieder, dass das Ergebnis der Anfrage alle Tupel enthält, die die Vermutung verletzen. Ist das Ergebnis leer, so gilt $DE \rightarrow B$.

```
select D, E
from R
group by D, E
  having count(distinct B) > 1;
```

## Hausaufgabe 2

Betrachten Sie das Relationenschema

PunkteListe: {Name, Aufgabe, Max, Erzielt, KlausurSumme, KNote, Bonus, GNote}

mit der folgenden beispielhaften Ausprägung:

| PunkteListe | | | | | | | |
|---|---|---|---|---|---|---|---|
| Name | Aufgabe | Max | Erzielt | KlausurSumme | KNote | Bonus | GNote |
| Bond | 1 | 10 | 4 | 18 | 2 | ja | 1.7 |
| Bond | 2 | 10 | 10 | 18 | 2 | ja | 1.7 |
| Bond | 3 | 11 | 4 | 18 | 2 | ja | 1.7 |
| Maier | 1 | 10 | 4 | 9 | 4 | nein | 4 |
| Maier | 2 | 10 | 2 | 9 | 4 | nein | 4 |
| Maier | 3 | 11 | 3 | 9 | 4 | nein | 4 |

1. Bestimmen Sie die geltenden FDs.
2. Bestimmen Sie die Kandidatenschlüssel.

1. Im Relationenschema gelten die folgenden funktionalen Abhängigkeiten:
   - {KNote, Bonus} $\rightarrow$ {GNote}
   - {Aufgabe} $\rightarrow$ {Max}
   - {KlausurSumme} $\rightarrow$ {KNote}
   - {Name, Aufgabe} $\rightarrow$ {Erzielt}
   - {Name} $\rightarrow$ {KlausurSumme, Bonus}

   Natürlich gelten auch alle anderen funktionalen Abhängigkeiten, die mit Hilfe der Armstrong-Axiome daraus hergeleitet werden können.
2. Der Kandidatenschlüssel ist {Name, Aufgabe}. Aus {Name} können die Attribute {KlausurSumme, Bonus}, aus {KlausurSumme} wiederum {KNote}, und aus {KNote, Bonus} dann {GNote} abgeleitet werden. Aus {Aufgabe} kann {Max} abgeleitet werden, und aus {Name, Aufgabe} noch das verbleibende Attribut {Erzielt}.

### Hausaufgabe 3

Betrachten Sie ein abstraktes Relationenschema $\mathcal{R} = \{A, B, C, D, E, F\}$ mit den FDs

- $A \to BC$
- $C \to DA$
- $E \to ABC$
- $F \to CD$
- $CD \to BEF$

1. Berechnen Sie die Attributhülle von $A$.
2. Bestimmen Sie alle Kandidatenschlüssel.
3. Bestimmen Sie zu den gebenen FDs die kanonische Überdeckung.
4. In welcher Normalform befindet sich die Relation?
5. Falls nötig, überführen Sie die Relation in die dritte Normalform, indem Sie den Synthesealgorithmus anwenden.

#### Attributhülle von A

Berechnung der Attributhülle von A mit Hilfe des bekannten *AttrHülle*-Algorithmus.

Aufruf: *AttrHülle*$(F, A)$.

| Schritt | betrachtete FD | Ergebnis |
|---|---|---|
| init | | $\{A\}$ |
| 1. | $A \to BC$ | $\{A, B, C\}$ |
| 2. | $C \to DA$ | $\{A, B, C, D\}$ |
| 3. | $CD \to BEF$ | $\{A, B, C, D, E, F\}$ |

Damit enthält die Attributhülle von $A$ alle Attribute.

#### Kandidatenschlüssel

$\{A\}$ ist nach der vorherigen Berechnung (Attributhülle von A) ein Superschlüssel. Da $\{A\}$ außerdem minimal ist, ist $\{A\}$ ein Kandidatenschlüssel. Da man aus $\{C\}$ und $\{E\}$ direkt $A$ folgern kann, handelt es sich hier ebenfalls um Superschlüssel und da sie einelementig sind (also minimal sind) auch um Kandidatenschlüssel. Aus $\{F\}$ wiederum kann $C$ und somit $A$ gefolgert werden. Damit ist $\{F\}$ analog zu oben auch ein Kandidatenschlüssel.

$\{B\}$ und $\{D\}$ sind dagegen keine Kandidatenschlüssel. $\{B\}$ ist nicht einmal Superschlüssel. $\{CD\}$ wäre zwar ein Superschlüssel, allerdings kein Kandidatenschlüssel, da nicht minimal.

Kandidatenschlüssel sind: $\{A\}, \{C\}, \{E\}, \{F\}$.

#### Kanonische Überdeckung

Sei FD die Menge der funktionalen Abhängigkeiten, also

$$FD = \{A \to BC, C \to DA, E \to ABC, F \to CD, CD \to BEF\}$$

**Kanonische Überdeckung**

Gegeben ist die Ausgangsmenge $FD$.

1. *Führe für jede FD* $\alpha \to \beta \in FD$ *die Linksreduktion durch.*

   Einzige in Betracht kommende FD ist $CD \to BEF$.

   - Ist $C$ überflüssig?
     $AttrHülle(F, D) = \{D\} \not\supseteq \{B, E, F\}$
   - Ist $D$ überflüssig?
     $AttrHülle(F, C) = \{A, B, C, D, E, F\} \supseteq \{B, E, F\}$

   Damit kann $CD \to BEF$ zu $C \to BEF$ reduziert werden.

2. *Führe für jede (verbliebene) FD* $\alpha \to \beta$ *die Rechtsreduktion durch.*

   Bisheriges Zwischenergebnis:

$$
\begin{aligned}
A &\to BC && (1)\\
C &\to DA && (2)\\
E &\to ABC && (3)\\
F &\to CD && (4)\\
C &\to BEF && (5)
\end{aligned}
$$

   **Betrachte FD (1):**

   - Ist $B$ überflüssig?
     $B \in AttrHülle(F - \text{FD } (1) \cup (A \to C), A)$, da $A \to C \to BEF$.
   - Ist $C$ überflüssig?
     $C \notin AttrHülle(F - \text{FD } (1) \cup (A \to \emptyset), A)$.

   Damit erhält man für FD (1): $A \to C$.

   **Betrachte FD (2):**

   - Ist $D$ überflüssig?
     $D \in AttrHülle(F - \text{FD } (2) \cup (C \to A), C)$, da $C \to BEF, F \to CD$.
   - Ist $A$ überflüssig?
     $A \in AttrHülle(F - \text{FD } (2) \cup (C \to \emptyset), C)$, da $C \to BEF, E \to ABC$.

   Damit erhält man für FD (2): $C \to \emptyset$.

   **Betrachte FD (3):**

   - Ist $A$ überflüssig?
     $A \notin AttrHülle(F - \text{FD } (3) \cup (E \to BC), E)$.
   - Ist $B$ überflüssig?
     $B \in AttrHülle(F - \text{FD } (3) \cup (E \to AC), E)$, da $E \to AC, C \to BEF$.
   - Ist $C$ überflüssig?
     $C \in AttrHülle(F - \text{FD } (3) \cup (E \to A), E)$, da $E \to A, A \to C$.

   Damit erhält man für FD (3): $E \to A$.

   **Betrachte FD (4):**

   - Ist $C$ überflüssig?
     $C \notin AttrHülle(F - \text{FD } (4) \cup (F \to D), F)$.

- Ist $D$ überflüssig?
  $D \notin \mathit{AttrHülle}(F - \text{FD } (4) \cup (F \to C), F)$.

  Damit bleibt FD (4) unverändert.

**Betrachte FD** (5)**:**

- Ist $B$ überflüssig?
  $B \notin \mathit{AttrHülle}(F - \text{FD } (5) \cup (C \to EF), C)$.
- Ist $E$ überflüssig?
  $E \notin \mathit{AttrHülle}(F - \text{FD } (5) \cup (C \to BF), C)$.
- Ist $F$ überflüssig?
  $F \notin \mathit{AttrHülle}(F - \text{FD } (5) \cup (C \to BE), C)$.

  Damit bleibt FD (5) unverändert.

3. *Entferne die FDs der Form* $\alpha \to \emptyset$.

   Bisheriges Zwischenergebnis:

$$\begin{aligned} A &\to C \\ C &\to \emptyset \qquad (6) \\ E &\to A \\ F &\to CD \\ C &\to BEF \end{aligned}$$

   FD (6) wird eliminiert.

4. *Fasse mittels der Vereinigungsregel FDs der Form* $\alpha \to \beta_1, \ldots, \alpha \to \beta_n$ *zusammen.*

   Bisheriges Zwischenergebnis:

$$F_c = \begin{cases} A \to C \\ E \to A \\ F \to CD \\ C \to BEF \end{cases}$$

   Es werden keine FDs vereinigt, da es keine zwei FDs mit gleicher linker Seite gibt.

   $F_c$ ist eine kanonische Überdeckung zur Ausgangsmenge $F$.

### Normalform?

Mindestens 3NF (um genau zu sein sogar 4NF).

### Dritte Normalform, wenn sie nicht schon in der 3NF wäre (laut Daniel Richter, danke!).

Bestimmen der kanonischen Überdeckung siehe oben. Für jede funktionale Abhängigkeit aus der kanonischen Überdeckung wird ein Relationenschema erstellt:

$$\begin{aligned} R_1 &: [A, C] \\ R_2 &: [E, A] \\ R_3 &: [F, C, D] \\ R_4 &: [C, B, E, F] \end{aligned}$$

R1 enthält einen der Kandidatenschlüssel (sogar zwei: nämlich $\{A\}$ und $\{C\}$), so dass kein zusätzliches Schema erstellt werden muss. Keines der Relationenschemata ist in einem anderen Schema enthalten, so dass nichts eliminiert werden kann.