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

Übung zur Vorlesung *Grundlagen: Datenbanken* im WS18/19

Moritz Sichert, Lukas Vogel (gdb@in.tum.de)
https://db.in.tum.de/teaching/ws1819/grundlagen/

Blatt Nr. 08

Tool zum Üben von funktionalen Abhängigkeiten: https://normalizer.db.in.tum.de/.

Hausaufgabe 1

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

1. $A \rightarrow BC$
2. $C \rightarrow DA$
3. $E \rightarrow ABC$
4. $F \rightarrow CD$
5. $CD \rightarrow BEF$

a) Berechnen Sie die Attributhülle von A.

b) Bestimmen Sie alle Kandidatenschlüssel.

c) Bestimmen Sie zu den gegebenen FDs die kanonische Überdeckung.

Lösung:

Attributhülle von A

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

Aufruf: *AttrHülle*$(FD, \{A\})$.

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

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

Gegeben ist die Ausgangsmenge $F = \{A \to BC, C \to DA, E \to ABC, F \to CD, CD \to BEF\}$.

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

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

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

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

     $\{C\}^+ = \{A, B, C, D, E, F\} \supseteq \{B, E, F\}$

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

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

   Betrachte FD (2):

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

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

   Betrachte FD (3):

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

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

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

Betrachte FD (4):

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

$$\begin{array}{lcl} A & \to & C \\ C & \to & \emptyset \\ E & \to & A \\ F & \to & CD \\ C & \to & BEF \end{array} \tag{6}$$

   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.

Hausaufgabe 2

Ist die kanonische Überdeckung F_c einer Menge F von funktionalen Abhängigkeiten eindeutig? Begründen Sie Ihre Antwort oder finden Sie ein Gegenbeispiel.

Lösung:

Die kanonische Überdeckung F_c zu einer Menge von funktionalen Abhängigkeiten F ist nicht eindeutig.

Begründung: Im Algorithmus zur Bestimmung der kanonischen Überdeckung ist nicht festgelegt, in welcher Reihenfolge die FDs bearbeitet werden.

Als Beispiel seien folgende funktionale Abhängigkeiten gegeben:

1. $A \rightarrow BC$
2. $B \rightarrow AC$

Wird die erste FD in der Rechtsreduktion zuerst abgearbeitet, ergibt sich:

$$F_c = \{A \rightarrow B, B \rightarrow AC\}$$

Wird die zweite FD in der Rechtsreduktion zuerst abgearbeitet, erhält man hingegen:

$$F_c = \{A \rightarrow BC, B \rightarrow A\}$$

Hausaufgabe 3

Geben Sie für jede der Normalformen 1NF, 2NF, 3NF, BCNF, 4NF jeweils eine Relation mit FDs an, sodass die Relation in der gewünschten Normalform ist (und in keiner höheren).

Lösung:

Für alle Normalformen betrachten wie die Relation $\mathcal{R} = \{A, B, C, D\}$.

- 1.NF:

  FDs:

  - $AB \rightarrow C$
  - $B \rightarrow D$

  Die Relation ist nicht mengenwertig, daher 1. NF. D ist lediglich von B abhängig, der Kandidatenschlüssel ist aber AB, weswegen D nicht voll funktional vom Kandidatenschlüssel abhängig ist, daher keine 2. NF.

- 2. NF:

  FDs:

  - $AB \rightarrow C$
  - $C \rightarrow D$

  Jedes Attribut der Relation ist voll funktional abhängig vom Kandidatenschlüssel AB, daher 2.NF. Das Attribut D ist transitiv und nicht direkt vom Kandidatenschlüssel abhängig, darum nicht 3. NF.

- 3. NF:

  FDs:

  - $AB \rightarrow CD$
  - $BC \rightarrow AD$
  - $D \rightarrow C$

Für alle FDs gilt entweder, dass sie trivial ist, dass die linke Seite Superschlüssel ist oder dass die rechte Seite in einem Kandidatenschlüssel enthalten ist, daher 3. NF. Bei der BCNF fällt die dritte erlaubte Art von FD weg, daher FDs müssen trivial sein oder ihre linke Seite Superschlüssel. Da die dritte FD des Beispiels dies verletzt ist die Relation nicht in BCNF und daher genau in 3. NF.

- BCNF:

  FDs:

  - $AB \rightarrow CD$
  - $BC \rightarrow AD$
  - $D \twoheadrightarrow C$

  BCNF, da die BCNF verletzende FD aus dem Beispiel für 3. NF entfernt wurde. Nicht 4. NF weil eine nicht trivial MVD gilt, deren linke Seite nicht Superschlüssel ist.

- 4. NF

  FDs:

  - $AB \rightarrow CD$
  - $BC \rightarrow AD$

  Nach Entfernung der nicht trivialen MVD dann auch 4. NF.

Hausaufgabe 4

Gegeben sei die durch folgende SQL-Statements definierte Ausprägung einer Relation.

```
create table kinder_fahrraeder (
  person varchar(100) not null,
  kind_name varchar(100) not null,
  kind_alter integer not null,
  fahrrad_typ varchar(100) not null,
  fahrrad_farbe varchar(100) not null
);
insert into kinder_fahrraeder values
  ('Thomas', 'Markus', 10, 'Trekking-Fahrrad', 'schwarz'),
  ('Thomas', 'Markus', 10, 'Mountainbike', 'rot'),
  ('Thomas', 'Johanna', 5, 'Trekking-Fahrrad', 'schwarz'),
  ('Thomas', 'Johanna', 5, 'Mountainbike', 'rot');
```

Es gelten die beiden komplementären MVD

1. person $\twoheadrightarrow$ {kind_name, kind_alter} und
2. person $\twoheadrightarrow$ {fahrrad_typ, fahrrad_farbe}

sowie die FD

3. kind_name $\rightarrow$ kind_alter.

a) Laura, das dritte Kind von Thomas, wird geboren. Fügen Sie Laura per SQL-Insert-Statement hinzu und beachten Sie dabei die MVDs. Formulieren Sie Ihr Statement so, dass es auch ohne Kenntnis der Fahrräder von Thomas funktionert (d.h. nicht `insert ... 'Mountainbike', 'rot');`).

b) Allgemein gesprochen: In eine Relation $R = \{A, B, C\}$ mit den MVDs $A \twoheadrightarrow B$ und $A \twoheadrightarrow C$ soll für ein bestimmtes a in Spalte A ein neuer Wert b in Spalte B eingefügt werden. Wie viele Tupel müssen hinzugefügt werden, damit die MVDs weiterhin gelten?

c) Was passiert, wenn Thomas seine beiden Fahrräder verkauft?

d) Überführen Sie die Relation `kinder_fahrraeder` mit dem Dekompositionsalgorithmus in die 4. NF.

e) Schreiben Sie ein SQL-Statement um zu prüfen ob die MVDs der Relation `kinder_fahrraeder` erfüllt sind.

Lösung:

a) Zunächst werden alle Fahrräder von Thomas ermittelt, anschließend wird Laura kombiniert mit jedem Fahrrad in die Tabelle eingefügt.

```
with fahrraeder as (
  select distinct person, fahrrad_typ, fahrrad_farbe
  from kinder_fahrraeder
  where person='Thomas'
)
insert into kinder_fahrraeder (
  select person, 'Laura', '0', fahrrad_typ, fahrrad_farbe
  from fahrraeder
);
```

b) b muss kombiniert mit jedem Wert in C eingefügt werden, die Anzahl neuer Tupel ist daher:

```
select count(distinct C) from R where A=a;
```

c) Da die fahrradbezogenen Spalten nicht nullable sind, müssen sämtliche Tupel gelöscht werden, sodass auch die Informationen zu den Kindern verloren gehen. Durch Normalisierung kann diese Anomalie verhindert werden.

d) Zunächst wird anhand der MVDs zerlegt, wir erhalten die Relationen:

$$\begin{aligned}\mathcal{R}_1 &= \{\underline{\text{person}, \text{kind\_name}}, \text{kind\_alter}\} \text{ und} \\ \mathcal{R}_2 &= \{\underline{\text{person}, \text{fahrrad\_typ}, \text{fahrrad\_farbe}}\}.\end{aligned}$$

Anschließend wird R_1 anhand der FD weiter zerlegt:

$$\begin{aligned}\mathcal{R}_{1a} &= \{\underline{\text{person}, \text{kind\_name}}\} \text{ und} \\ \mathcal{R}_{1b} &= \{\underline{\text{kind\_name}}, \text{kind\_alter}\} \text{ und} \\ \mathcal{R}_2 &= \{\underline{\text{person}, \text{fahrrad\_typ}, \text{fahrrad\_farbe}}\}.\end{aligned}$$

Alternativ kann auch mit der Zerlegung anhand der FD begonnen werden, die finale Zerlegung ist dieselbe.

e) Eine Möglichkeit ist es, das Kreuzprodukt aus allen Fahrrädern und Kindern pro Person zu bilden, und die tatsächliche Ausprägung abzuziehen. Ist das Ergebnis des SQL-Statements leer, so sind die MVDs erfüllt.

```
with kinder as (
  select distinct person, kind_name, kind_alter
  from kinder_fahrraeder
), fahrraeder as (
  select distinct person, fahrrad_typ, fahrrad_farbe
  from kinder_fahrraeder
)
select f.person, k.kind_name, k.kind_alter, f.fahrrad_typ, f.fahrrad_farbe
from fahrraeder f, kinder k
where f.person=k.person
except
select * from kinder_fahrraeder;
```