



Übung zur Vorlesung *Grundlagen: Datenbanken* im WS20/21
Christoph Anneser, Josef Schmeißer, Moritz Sichert, Lukas Vogel (gdb@in.tum.de)
<https://db.in.tum.de/teaching/ws2021/grundlagen/>

Blatt Nr. 07

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.
- (d) Überführen Sie die Relation in die dritte Normalform, indem Sie den Synthesealgorithmus anwenden.

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 $F = \{A \rightarrow BC, C \rightarrow DA, E \rightarrow ABC, F \rightarrow CD, CD \rightarrow BEF\}$.

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

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

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

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

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

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

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

Bisheriges Zwischenergebnis:

$$A \rightarrow BC \tag{1}$$

$$C \rightarrow DA \tag{2}$$

$$E \rightarrow ABC \tag{3}$$

$$F \rightarrow CD \tag{4}$$

$$C \rightarrow BEF \tag{5}$$

Betrachte FD (1):

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

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

Betrachte FD (2):

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

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

Betrachte FD (3):

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

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

Betrachte FD (4):

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

Damit bleibt FD (4) unverändert.

Betrachte FD (5):

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

Damit bleibt FD (5) unverändert.

3. Entferne die FDs der Form $\alpha \rightarrow \emptyset$.

Bisheriges Zwischenergebnis:

$$\begin{array}{l}
 A \rightarrow C \\
 C \rightarrow \emptyset \\
 E \rightarrow A \\
 F \rightarrow CD \\
 C \rightarrow BEF
 \end{array} \tag{6}$$

FD (6) wird eliminiert.

4. Fasse mittels der Vereinigungsregel FDs der Form $\alpha \rightarrow \beta_1, \dots, \alpha \rightarrow \beta_n$ zusammen.

Bisheriges Zwischenergebnis:

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

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

F_c ist eine kanonische Überdeckung zur Ausgangsmenge F .

Dritte Normalform

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

$$\begin{aligned} R_1 &= \{\underline{A}, C\} \\ R_2 &= \{\underline{E}, A\} \\ R_3 &= \{\underline{F}, C, D\} \\ R_4 &= \{\underline{C}, B, E, F\} \end{aligned}$$

R_1 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.

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

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

sowie die FD

3. $\text{kind_name} \rightarrow \text{kind_alter}$.

- 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 funktioniert (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:

$$R_1 : \{[\underline{\text{person}}, \text{kind_name}, \text{kind_alter}]\} \text{ und}$$

$$R_2 : \{[\underline{\text{person}}, \text{fahrrad_typ}, \text{fahrrad_farbe}]\}.$$

Anschließend wird R_1 anhand der FD weiter zerlegt:

$$R_{1a} : \{[\underline{\text{person}}, \text{kind_name}]\} \text{ und}$$

$$R_{1b} : \{[\underline{\text{kind_name}}, \text{kind_alter}]\} \text{ und}$$

$$R_2 : \{[\underline{\text{person}}, \text{fahrrad_typ}, \text{fahrrad_farbe}]\}.$$

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;
```

Hausaufgabe 4

Überführen Sie verlustlos in die 4. NF:

$$\mathcal{R} = \{A, B, C, D\} \text{ mit den FDs } F = \{C \rightarrow D, AB \rightarrow D, DA \rightarrow B, BA \rightarrow C\}.$$

Lösung:

Es kann entweder der BCNF- oder der 4NF-Algorithmus angewendet werden. Im ersten Fall erfolgt die Zerlegung wie zuvor geübt und das Ergebnis ist in BCNF. Da keine MVD gilt, die nicht aus FDs hergeleitet werden kann, ist die Relation damit automatisch auch in 4NF. Im anderen Fall gilt, dass alle FDs auch MVDs sind und man zieht diese für die Zerlegung heran.

Ein mögliches Ergebnis wäre:

$$\mathcal{R}_1 = \{\underline{C}, D\}$$

$$\mathcal{R}_2 = \{\underline{A}, \underline{B}, C\}$$