



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. 07

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

Hausaufgabe 1

Gegeben sei eine Relation

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

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

- AB ist ein Superschlüssel der Relation
- $DE \rightarrow B$

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

Lösung:

- Durch Gruppierung nach A und B kann 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 jedoch leer, so ist AB ein Superschlüssel.

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

- 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

Fahrplan: {[Linie, Verband, von, nach, von GPS, nach GPS, Preis, #Fahrzeuge, Modus]}

mit der folgenden beispielhaften Ausprägung:

Linie	Verbund	von	nach	von GPS	nach GPS	Preis	#Fahrzeuge	Modus
U6	MVV	GF	G	0N 0W	1S 0W	1€	20	U-Bahn
U6	MVV	G	GH	1S 0W	2S 0W	1€	20	U-Bahn
U6	MVV	GH	FR	2S 0W	5S 0W	3€	20	U-Bahn
U3	MVV	MF	GI	8S 0W	9S 0W	1€	16	U-Bahn
690	MVV	GF	DI	0N 0W	1N 0W	1€	5	Bus
690	MVV	DI	NF	1N 0W	3N 1W	2€	5	Bus
690	MVV	NF	EH	3N 1W	5N 2W	2€	5	Bus
S1	MVV	NF	EH	3N 1W	5N 2W	3€	8	S-Bahn

- a) Bestimmen Sie die geltenden FDs.
b) Bestimmen Sie die Kandidatenschlüssel.

Lösung:

- a) Im Relationenschema gelten die folgenden funktionalen Abhängigkeiten:

- $\{\text{Linie}\} \rightarrow \{\#\text{Fahrzeuge}, \text{Modus}\}$
- $\{\text{von}\} \rightarrow \{\text{von GPS}\}$
- $\{\text{nach}\} \rightarrow \{\text{nach GPS}\}$
- $\{\text{von GPS}\} \rightarrow \{\text{von}\}$
- $\{\text{nach GPS}\} \rightarrow \{\text{nach}\}$
- $\{\text{Linie}, \text{von}, \text{nach}\} \rightarrow \{\text{Preis}\}$
- $\emptyset \rightarrow \{\text{Verbund}\}$

Natürlich gelten auch alle anderen funktionalen Abhängigkeiten, die mit Hilfe der Armstrong-Axiome daraus hergeleitet werden können.

- b) Die möglichen Kandidatenschlüssel sind:

- $\{\text{Linie}, \text{von}, \text{nach}\}$
- $\{\text{Linie}, \text{von GPS}, \text{nach}\}$
- $\{\text{Linie}, \text{von}, \text{nach GPS}\}$
- $\{\text{Linie}, \text{von GPS}, \text{nach GPS}\}$

Aus „Linie“ können die Attribute „#Fahrzeuge“ und „Modus“, aus „Linie“, „von“ und „nach“ das Attribut „Preis“ abgeleitet werden. Da „von“ und „von GPS“ sowie „nach“ und „nach GPS“ jeweils in beide Richtungen abgeleitet werden können, enthält der Kandidatenschlüssel alle möglichen Kombinationen.

Hausaufgabe 3

Bringen Sie die folgende Relation verlustlos und abhängigkeitsbewahrend in die 3. NF.

$$R : \{[A, B, C, D, E, F]\}$$

FDs:

1. $AB \rightarrow CD$
2. $ABC \rightarrow D$
3. $E \rightarrow C$
4. $D \rightarrow C$
5. $CDE \rightarrow AB$

Beachten Sie, dass es für die Lösung notwendig ist, einen Kandidatenschlüssel zu ermitteln, jedoch nicht alle Kandidatenschlüssel. Beachten Sie außerdem, dass die Relation das Attribut F enthält, welches bei der Zerlegung nicht wegfallen darf.

Lösung:

Der Kandidatenschlüssel **muss** bestimmt werden, anders verliert man im dritten Schritt des Synthesalgorithmus potentiell Semantik und in diesem Fall insbesondere das Attribut F . Es müssen (da nicht explizit gefordert) jedoch nicht alle Kandidatenschlüssel sondern lediglich einer bestimmt werden. Ideales Verfahren hierzu ist, dass wir betrachten, welche Attribute nicht auf der rechten Seite von mindestens einer FD stehen. Diese **müssen** im Kandidatenschlüssel enthalten sein. Dies ist hier für E und F der Fall. Nun untersuchen wir, ob EF bereits ein Schlüssel für die Relation ist, womit wir dann bereits fertig wären. Wegen $AttrHuelle(FD, \{E, F\}) = \{E, F, C\} \neq R$ ist dies nicht der Fall. Wir versuchen nun möglichst intelligent ein Attribut hinzuzufügen, um Schlüsseleigenschaft zu erreichen. Da aus DE sofort AB folgt und E bereits definitiv im Schlüssel enthalten sein muss, fügen wir D hinzu. Damit gilt $AttrHuelle(FD, \{E, F, D\}) = \{E, F, D, C, A, B\} = R$, daher EDF ist Schlüssel. Insbesondere ist EDF auch Kandidatenschlüssel, da EF wie zuvor erklärt in jedem Schlüssel enthalten sein müssen, jedoch alleine kein Schlüssel sind, Reduktion nicht möglich, Schlüssel ist also Kandidatenschlüssel.

Ein weiterer Kandidatenschlüssel ist $ABEF$.

Nach dieser kurzen Überlegung muss lediglich der Synthesalgorithmus ausgeführt werden. Der Algorithmus ist in den Folien, im Buch und auf Wikipedia im Detail beschrieben. Die genaue Herleitung können Sie sich mit dem eingangs genannten FD-Tool generieren lassen, hier lediglich das Ergebnis zum Abgleich:

Kanonische Überdeckung:

- $AB \rightarrow D$
- $E \rightarrow C$
- $D \rightarrow C$
- $DE \rightarrow AB$

Entstehende Relationen also:

- $R_1(A, B, D, E)$ mit zugeordneter FD $AB \rightarrow D$ sowie $DE \rightarrow AB$ (Zusammenfassung der Relationen, da eine in der anderen enthalten.)

- $R_2(E, C)$ mit zugeordneter FD $E \rightarrow C$
- $R_3(D, C)$ mit zugeordneter FD $D \rightarrow C$
- $R_4(E, F, D)$ ohne FD, für gewählten Kandidatenschlüssel eingefügt, da Kandidatenschlüssel nirgendwo enthalten.

Hausaufgabe 4

Bestimmen Sie alle Kandidatenschlüssel der Relation R . Wenden Sie den Dekompositionsalgorithmus an, um die Relation R in die BCNF zu zerlegen und unterstreichen Sie die Schlüssel der Teilrelationen des Endergebnisses.

$$R : \{[A, B, C, D, E, F]\}$$

FDs:

1. $B \rightarrow DA$
2. $DEF \rightarrow B$
3. $C \rightarrow EA$

Prüfen Sie als erstes, ob FD 1) für die Zerlegung geeignet ist und - falls dies der Fall ist - verwenden Sie diese im ersten Zerlegungsschritt. Für diese Aufgabe ist zu bedenken, dass die oben angegebenen FDs eine Charakterisierung der insgesamt geltenden FDs sind. Die Menge der geltenden FDs ist größer. Wieso? Wie muss dies beim Dekompositionsalgorithmus genutzt werden?

Lösung:

- Dekompositionsalgorithmus:
 - Starte mit $Z := \{R\}$.
 - R in BCNF? - Nein, $B \rightarrow DA$ verletzt die BCNF.
 - * Zerlegung anhand FD $B \rightarrow DA$, da $\{B\}$ kein Superschlüssel:
 $R_1 = \{A, B, D\}$ mit den FDs $F_1 = \{B \rightarrow DA\}$,
 $R_2 = \{B, C, E, F\}$ mit den FDs $F_2 = \{C \rightarrow E\}$, FD (2) geht verloren und FD (3) geht "teilweise" verloren: wenn $C \rightarrow AE$ gilt, dann gilt auch $C \rightarrow A$ und $C \rightarrow E$ (Dekompositionsregel), aber lediglich $C \rightarrow E$ bleibt erhalten.
 $Z := \{R_1, R_2\}$
 - R_1 in BCNF? - Ja.
 - R_2 in BCNF? - Nein, $C \rightarrow E$ verletzt die BCNF.
 - * Zerlegung anhand FD $C \rightarrow E$, da $\{C\}$ kein Superschlüssel:
 $R_{2,1} = \{C, E\}$ mit den FDs $F_{2,1} = \{C \rightarrow E\}$,
 $R_{2,2} = \{B, C, F\}$ mit ausschließlich trivialen FDs.
 $Z := \{R_1, R_{2,1}, R_{2,2}\}$
 - $R_{2,1}$ in BCNF? - Ja.
 - $R_{2,2}$ in BCNF? - Ja.

- Ergebnis:

$$\begin{aligned}R_1 &= \{A, \underline{B}, D\} \\R_{2.1} &= \{\underline{C}, E\} \\R_{2.2} &= \{\underline{B}, \underline{C}, \underline{F}\}\end{aligned}$$

Im Allgemeinen ist eine gegebene FD-Menge weder minimal noch vollständig. Die angegebenen FDs enthalten also möglicherweise Redundanzen einerseits und andererseits werden triviale Abhängigkeiten i.d.R. nicht explizit mit angegeben. Bei der Ausführung des Dekompositionsalgorithmus müssen jedoch alle *geltenden* FDs betrachtet werden, die sich mit Hilfe der Axiome von Armstrong herleiten lassen (F^+). So gilt im obigen Beispiel in R_2 die FD $C \rightarrow E$, obwohl diese nicht explizit angegeben war.