Übung zur Vorlesung *Grundlagen: Datenbanken* im WS14/15

Harald Lang (harald.lang@in.tum.de)
http://www-db.in.tum.de/teaching/ws1415/grundlagen/

Blatt Nr. 7

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.

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:

$$A \to BC \quad (1)$$
$$C \to DA \quad (2)$$
$$E \to ABC \quad (3)$$
$$F \to CD \quad (4)$$
$$C \to BEF \quad (5)$$

   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:

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

   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.

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.

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

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 FD 1) ob Sie 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?

- 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.