

Introduction to Data and Knowledge Engineering

4. Übung

Syntheseverfahren



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Aufgabe 3.3

FD-Äquivalenzklassen

Bestimmen Sie die Äquivalenzklassen für die FD-Menge F

$$F = \{B \rightarrow C, A \rightarrow DE, B \rightarrow F, F \rightarrow GH, D \rightarrow I\}$$

Sei $[X_1..X_n]$ die Menge der Attribute die äquivalent zu $X_1..X_n$ sind.

$$[B] = \{B\}, [A] = \{A\}, [F] = \{F\}, [D] = \{D\}$$

Seien $[[X_1..X_n \rightarrow Y_1..Y_m]]$ die FDs, deren linke Seite äquivalent zu $X_1..X_n$ sind.

Äquivalenzklassen	Teilschemata	Schlüsselkandidaten
$[[B \rightarrow C]] = \{B \rightarrow C, B \rightarrow F\}$	(BCF)	{B}
$[[A \rightarrow DE]] = \{A \rightarrow DE\}$	(ADE)	{A}
$[[F \rightarrow GH]] = \{F \rightarrow GH\}$	(FGH)	{F}
$[[D \rightarrow I]] = \{D \rightarrow I\}$	(DI)	{D}

Aufgabe 4.1

Verbundtreue einer Zerlegung

Im Text der Aufgabe 4.1 steht, dass ein Zerlegung eines Schemas verbundstreu ist, wenn sich die aus Teilschemen mittels natural join über Schlüsselkandidaten das originale Schema wieder zusammenfügen lässt.

- hier wird join auf Schemata mit Schlüsselkandidaten angewendet
- wenn die Bedingung erfüllt ist, dann kann man auch verlangen, dass der join immer über die Schlüsselkandidaten die linken Schemas erfolgt
- damit steht *hier* schon fest welche Schema mit welche gejoint wird, da für die linken Seiten jeweils nur ein rechter Partner in Frage kommt

$$R2:(ADE,\{A\}) \triangleright\triangleleft R4:(DI, \{D\}) = R24:(ADEI,\{A\}),$$

$$R1:(BCF,\{B\}) \triangleright\triangleleft R3:(FGH,\{F\}) = R13:(BCFGH,\{B\})$$

- R24 und R13 haben gar kein gemeinsames Attribut. Die Bedingung ist also nicht erfüllt.

Aufgabe 4.1

Syntheseverfahren mit Dummy-FD

$F' = \{B \rightarrow C, A \rightarrow DE, BC \rightarrow F, F \rightarrow GH, A \rightarrow I, D \rightarrow I, ABCDEFGHI \rightarrow \delta\}$

1. überflüssige Attribute: C in $BC \rightarrow F$ und CDEFGHI in $ABCDEFGHI \rightarrow \delta$

2. überflüssige FDs: $A \rightarrow I$

3. FDs mit identischer linker Seite: $B \rightarrow C, B \rightarrow F$

4. Äquivalente linke Seiten: keine

5. Äquivalenzklassen Teilschemata mit Schlüsselkandidaten

$\{B \rightarrow C, B \rightarrow F\}$ R1: (BCF, {B})

$\{A \rightarrow DE\}$ R2: (ADE, {A})

$\{F \rightarrow GH\}$ R3: (FGH, {F})

$\{D \rightarrow I\}$ R4: (DI, {D})

$\{AB \rightarrow \delta\}$ R5: (AB, {AB})

Aufgabe 4.1

Syntheseverfahren mit Dummy-FD

Das Hinzunehmen der dummy-FD uns ein weiteres Teilschema $R5:(AB,\{AB\})$ geliefert. Damit erhalten wir die Verbundtreue.

Wie gehabt können auf die erforderliche Weise

$R24:(ADEI,\{A\})$ und $R13:(BCFGH,\{B\})$

durch join bilden und darüber hinaus

$R5:(AB,\{AB\}) \triangleright \triangleleft R24:(ADEI,\{A\}) = R524:(ABDEI,\{AB\})$ sowie

$R524:(ABDEI,\{AB\}) \triangleright \triangleleft R13:(BCFGH,\{B\}) = R52413:(ABCDEFGHI,\{AB\})$

und damit das Schema von dem wir ausgegangen waren.

Man kann sich das dummy-Attribut δ auch als einen versteckten Schlüssel (z.B. Index des jeweiligen Tupels) vorstellen.

Aufgabe 4.2

Schlüsselkandidaten

R:(ABCDEFGG) mit $F=\{A\rightarrow C, B\rightarrow DE, C\rightarrow F, F\rightarrow G, FG\rightarrow A, E\rightarrow G, ABCDEFG \rightarrow \delta\}$

Schlüsselkandidaten: Bilde zunächst maximale rechte Seite für linke Seite.

$A\rightarrow C, C\rightarrow F, F\rightarrow G : A\rightarrow ACFG$

$B\rightarrow DE, E\rightarrow G : B\rightarrow BDEG$

$C\rightarrow F, F\rightarrow G, FG\rightarrow A : C\rightarrow ACFG$

$F\rightarrow G, FG\rightarrow A, A\rightarrow C : F\rightarrow ACFG$

$FG\rightarrow A, A\rightarrow C : FG\rightarrow ACFG$

$E\rightarrow G : E\rightarrow EG$

Superschlüssel ergeben sie daraus viele, aber nur die folgenden sind minimal und damit Schlüsselkandidaten: { AB, BC, BF }

Aufgabe 4.2

Vorliegende NF

R:(ABCDEFGG) mit $F=\{A\rightarrow C, B\rightarrow DE, C\rightarrow F, F\rightarrow G, FG\rightarrow A, E\rightarrow G, ABCDEFG \rightarrow \delta\}$
Schlüsselkandidaten: { AB, BC, BF }

Da wir keine Primärschlüssel aus den Schlüsselkandidaten ausgesucht haben müssen wir alle Kandidaten ausprobieren und prüfen, ob wenigsten für einen die NF vorliegt (verallgemeinerte Definition der NF):

- 1NF: ist klar
- 2NF: für alle Schlüsselkandidaten ist DE nicht vollständig von dem Schlüssel abhängig (sondern nur von B). D.h. 2NF ist nicht erfüllt und damit auch nicht 3NF.

Aufgabe 4.2

Vereinfachtes Syntheseverfahren

R:(ABCDEFGG) mit $F=\{A\rightarrow C, B\rightarrow DE, C\rightarrow F, F\rightarrow G, FG\rightarrow A, E\rightarrow G, ABCDEFG\rightarrow\delta\}$

1. überflüssige Attribute:

G in $FG\rightarrow A$

CDEFG in $ABCDEFG \rightarrow \delta$

(mehrere Möglichkeiten: statt AB können auch CB bzw. FB übrig bleiben)

2. überflüssige FD: keine

3. FDs mit identischer linken Seite: $F\rightarrow G$ und $F\rightarrow A$ haben identische linke Seite

Aufgabe 4.2

Vereinfachtes Syntheseverfahren

4. äquivalente linke Seiten: A, C und F sind äquivalent

$$A \rightarrow C, C \rightarrow F, F \rightarrow A$$

5. Äquivalenzklassen

$$\{ A \rightarrow C, C \rightarrow F, F \rightarrow A, F \rightarrow G \}$$

$$\{ B \rightarrow DE \}$$

$$\{ E \rightarrow G \}$$

$$\{ AB \rightarrow \delta \}$$

Teilschemata mit Schlüsselkandidaten

$$R1: (ACFG, \{A, C, F\})$$

$$R2: (BDE, \{B\})$$

$$R3: (EG, \{E\})$$

$$R4: (AB, \{AB\})$$

Statt $R4: (AB, \{AB\})$ wird auch $R4: (AB, \{ \{AB\} \})$ geschrieben:

Die Menge der Schlüsselkandidaten enthält einen Schlüsselkandidate, der eine Menge ist mit den beiden Attributen A und B.

Entsprechend wird aus $R1: (ACFG, \{A, C, F\})$ dann $R1: (ACFG, \{\{A\}, \{C\}, \{F\}\})$

Die Schreibweise mit nur einer Klammerung passt besser zu der Schreibweise der FDs.

Der Begriff „Relation“ wird oft verschiedene Weise verwendet.

Inhalt der Tabelle, „Menge von Tupeln“

Eine Relationen-Schema (mit oder ohne Schlüssel)

Eine Relationen-Schema mit FD-Menge

Aufgabe 4.3

Syntheseverfahren mit Dummy-FD

$F = \{AB \rightarrow CD, DE \rightarrow AB, AC \rightarrow F, BF \rightarrow EG, E \rightarrow C, G \rightarrow H, H \rightarrow BF, ABCDEFGH \rightarrow \delta\}$

1. überflüssige Attribute Entfernen: CDEFGH in $ABCDEFGH \rightarrow \delta$

Das ist nur eine von mehreren Möglichkeiten. Generell ist klar: Wenn die linke Seite ein Schlüsselkandidat enthält können wir alle anderen Attribute aus die linken Seite entfernen (und auf der rechten Seite alle Attribute hinzufügen, aber wollen ja weniger Attribute nicht mehr davon).

Menge der Schlüsselkandidaten: $\{AB, DE, AG, AH, BDF, DG, DH\}$

2. überflüssige FD entfernen: keine FD ist überflüssig

3. FD mit identischer linken Seite: $AB \rightarrow CD$ und $AB \rightarrow \delta$

Aufgabe 4.3

Syntheseverfahren mit Dummy-FD

4. äquivalente linke Seiten:

- $AB \rightarrow DE$: folgt aus $AB \rightarrow CD$, $AC \rightarrow F$ und $BF \rightarrow EG$

$DE \rightarrow AB$

- $BF \rightarrow G$ folgt aus $BF \rightarrow EG$

$G \rightarrow H$

$H \rightarrow BF$

- | | |
|---|--------------------------------------|
| 5. Äquivalenzklassen | Teilschemata mit Schlüsselkandidaten |
| $\{ AB \rightarrow DE, AB \rightarrow \delta, DE \rightarrow AB \}$ | R1: (ABDCE, {AB, DE}) |
| $\{ AC \rightarrow F \}$ | R2: (ACF, {AC}) |
| $\{ BF \rightarrow EG, G \rightarrow H, H \rightarrow BF \}$ | R3: (BEFGH, {BF, G, H}) |
| $\{ E \rightarrow C \}$ | R4: (EC, {E}) |

Aufgabe 4.3

Erfüllte Normalformen

Welche Normalformen werden von allen Schemata und allen Schlüsselkandidaten erfüllt?

1NF: Potz-Blitz.

2NF: Nein. R1 mit Primärschlüssel DE erfüllt 2NF nicht, da $E \rightarrow C$ gilt

Wir können C aber R1 entfernen, da in R4 C als Extra-Attribut vorkommt.

Das Vollständige Syntheseverfahren enthält auch diesen Schritt.

Danach ist das Datenbank-Schema mit der FD-Menge F in 3NF