

Introduction to Data and Knowledge Engineering Sommersemester 2010



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Übung 4: Syntheseverfahren

21. Mai 2010

Aufgabe 4.1 Syntheseverfahren Verstehen

Das *Syntheseverfahren von Bernstein* ist ein Algorithmus, mit dem wir, ausgehend von einem flachen Relationen-Schema ohne Schlüssel-Definition (d.h. einer Menge von Attributen) und einer Menge von FDs auf diesen Attributen, eine Datenbank-Schema inklusive Schlüssel gewinnen, welches zu wenig Anomalien führt.

Das *Vollständige Syntheseverfahren* garantiert ein Datenbank-Schema in 3NF. Wir werden hier eine Variante behandeln, das *Vereinfachte Syntheseverfahren*, welche diese Garantie nicht leistet, die aber mit nicht zu viel Aufwand in der Übung ausprobiert werden kann.

Das Vereinfachte Syntheseverfahren besteht aus den folgenden Schritten:

- entferne überflüssige Attribute der FDs
- entfernen überflüssige FDs
- gruppiere die FDs zu Äquivalenzklassen bezüglich der linken Seiten und bilde daraus die Teil-Schemata.

Jede FD-Äquivalenzklasse bildet eine Teil-Schema, das alle Attribute enthält, die in der jeweiligen FD-Teilmenge vorkommen, und für die die linke Seiten der Äquivalenzklasse die Schlüsselkandidaten sind. Das Ergebnis ist also eine Datenbank-Schema, bei dem wir aus den jeweiligen Schlüsselkandidaten noch jeweils einen Primär-Schlüssel aussuchen müssen. In vorangegangenen Aufgaben haben wir die Schritte schon mal durchgeführt. Ausgehend von dem Schema

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

haben wir in Aufgabe 2.3 überflüssige Attribute und FDs entfernt und die folgende reduzierte aber äquivalente FD-Menge erhalten.

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

Zu dieser FD-Menge haben wir dann in Aufgabe 3.3 die Äquivalenzklassen inklusive eindeutiger Schlüsselkandidaten erhalten.

Äquivalenzklassen	Teil-Schemata	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)	$\{\{D\}\}$

Eine Zerlegung eines Relationen-Schemas in mehrere Teilschemata heißt *verbundtreu*, wenn die Zusammenfügung der Teil-Tabellen durch natural join die Tabelle ergibt, wie sie vor der Zerlegung war. Dies ist der Fall, wenn sich das Gesamt-Schema durch iterative natural joins der Teil-Schemata wiederherstellen lässt, wobei beim jedem join die übereinstimmenden Attribute ein Schlüsselkandidat eines der beiden Teil-Schemata sind.

Ist die oben besprochene Zerlegung verbundtreu? Wenn ja, schlagen Sie einen Purzelbaum. Falls nein, wenden Sie folgende Variante des Verfahrens an, welches die Verbundtreue garantiert. Das *Vereinfachten Syntheseverfahren mit dummy-FD* arbeitet wie das Vereinfachten Syntheseverfahren mit dem Unterschied, dass wir zur ursprünglichen FD-Menge die *dummy-FD* anfügen. Diese besteht auf der linken Seite aus allen Attributen, die in der FD-Menge vorkommen und auf der rechten aus einem dummy-Attribut, das keinem anderen Attribut entspricht und als kleines Delta geschrieben wird.

$$F \mapsto F \cup \{ABCDEF \rightarrow \delta\}$$

Wenden Sie das Vereinfachten Syntheseverfahren mit dummy-FD auf F an und zeigen Sie, dass das Ergebnis verbundtreu ist.

Aufgabe 4.2 Syntheseverfahren Anwenden 1

Sei $R = ABCDEFG$ ein Relationen-Schema und $F = \{A \rightarrow C, B \rightarrow DE, C \rightarrow F, F \rightarrow G, FG \rightarrow A, E \rightarrow G\}$ eine Menge FDs zu diesem Schema. Bestimmen Sie die Schlüsselkandidaten, die höchste vorliegende Normalform (bis 3NF) und wenden Sie gegebenenfalls das vereinfachte Syntheseverfahren mit dummy-FD an.

Aufgabe 4.3 Syntheseverfahren Anwenden 2

Sei $F = \{AB \rightarrow CD, DE \rightarrow AB, AC \rightarrow F, BF \rightarrow EG, E \rightarrow C, G \rightarrow H, H \rightarrow BF\}$ eine Menge von FDs für Relationen-Schema $R = (ABCDEFGH)$ ein. Wenden Sie das vereinfachte Syntheseverfahren mit dummy-FD an. Überprüfen Sie, ob das Ergebnis die 3NF erfüllt. Falls nein, wie müssen die Teil-Schemata verändert werden um 3NF zu erfüllen?