

Normalisierung (korrigiert)

Korrektur in Aufgabe 10.2 fett gedruckt. Wir verlangen natürlich nicht, dass alle 708 nicht-trivialen links-irreduziblen funktionalen Abhängigkeiten niedergeschrieben oder gar gedruckt werden. Der Aufgabe fehlte der Zusatz, dass die gesuchten Abhängigkeiten noch rechts-maximal sein sollen. Dieser Zusatz führt zu einer handhabbaren Menge an FDs und Schlüsselkandidaten sind notwendigerweise automatisch ablesbar.

Aufgabe 10.1 (3 Punkte): Erläutern Sie zwei Argumente, wofür normalisierte Tabellen nützlich sind. Beschreiben Sie eine Situation, in der eher mit denormalisierten Tabellen gearbeitet werden sollte.

Aufgabe 10.2 (5 Punkte): Auf der Website zur Vorlesung befindet sich eine TSV-Datei mit einem Exzerpt der IMDB (Internet Movie Database). In dieser Datei stellt die erste Zeile (der Header) die Attributnamen dar.

Finden Sie alle nicht-trivialen links-irreduziblen **rechts-maximalen**¹ funktionalen Abhängigkeiten, die von der angegebenen Instanz erfüllt werden. Sind all diese Abhängigkeiten solche, die der Semantik der Domäne (Film) folgen? Welche Attributmengen sind Kandidatenschlüssel?

Aufgabe 10.3 (7 Punkte): Gegeben sei nun die folgende Menge funktio-

¹Eine funktionale Abhängigkeit $X \rightarrow Y$ ist *rechts-maximal*, falls es keine gültige funktionale Abhängigkeit $X \rightarrow Y'$ gibt, für die $Y \subsetneq Y'$. Anders ausgedrückt, muss für alle Abhängigkeiten $X \rightarrow Y'$ gelten, dass $Y' \subseteq Y$. Y kann also nicht mehr vergrößert werden.

naler Abhängigkeiten² des Relationsschemas $R(A, B, C, D, E, F, G, H)$.

$$F = \left\{ \begin{array}{l} AE \rightarrow B, \quad BC \rightarrow F, \quad F \rightarrow G, \\ G \rightarrow H, \quad DH \rightarrow E, \quad ACD \rightarrow H \end{array} \right\}$$

1. Ist die Menge minimal? Falls ja, begründen Sie Ihre Antwort. Falls nicht, minimieren Sie die Menge der funktionalen Abhängigkeiten.
2. Bestimmen Sie alle Schlüsselkandidaten (candidate keys) von R gemäß den gegebenen funktionalen Abhängigkeiten.
3. Ist das Schema in der dritten Normalform? Falls nicht, überführen Sie das Modell in eines, sodass alle Schemata die dritte Normalform erfüllen, aber ohne Verluste (*lossless* bzw. verbundtreu). Ist Ihre Zerlegung auch abhängigkeittreu³?

Aufgabe 10.4 (5 Punkte): Arbeiten Sie für diese Aufgabe mit dem Beispiel zur BCNF aus den Vorlesungsfolien. Zur Erinnerung, das Schema welches in BCNF zerlegt werden soll ist $R(\text{student}, \text{topic}, \text{advisor})$ mit den funktionalen Abhängigkeiten $\{\text{student}, \text{topic}\} \rightarrow \text{advisor}$ und $\text{advisor} \rightarrow \text{student}$.

1. Warum ist R in der dritten Normalform, nicht aber in der BCNF?
2. Betrachten Sie nun die drei vorgeschlagenen Zerlegungen auf den Folien, also
 - (a) $R_1(\text{student}, \text{topic})$ und $R_2(\text{student}, \text{advisor})$,
 - (b) $R_1(\text{topic}, \text{student})$ und $R_2(\text{topic}, \text{advisor})$, sowie
 - (c) $R_1(\text{advisor}, \text{topic})$ und $R_2(\text{advisor}, \text{student})$.

Entscheiden Sie für jede der drei Zerlegungen, ob sie verbundtreu (*lossless*) ist. Begründen Sie Ihre jeweilige Antwort.

Aufgabe 10.5 (5 Punkte): Stellen Sie sich folgende Normalform eines Datenbankschemas vor: Jede Relation ist binär und benutzt als Schlüssel beide Attribute. Ordnen Sie diese Normalform in die bestehenden Normalformen ein: Ist sie restriktiver als die 3NF? Ist diese Normalform wünschenswert?

²Als Kurzschreibweise für Attributmengen benutzen wir die Konkatenationsschreibweise, also wird $\{A, B, E\}$ durch ABE notiert.

³Abhängigkeitstreue heißt, dass jede ursprüngliche funktionale Abhängigkeit auch in der Zerlegung Bestand hat, d. h. es gibt irgendeine Tabelle/ein Relationsschema, in der die funktionale Abhängigkeit weiterhin realisiert ist.

Welche Vor-/Nachteile sehen Sie für relationale Datenbanken in der erdachten Normalform voraus?

Betrachten Sie nun folgende Normalform unter denselben Gesichtspunkten wie eben: Jede Relation hat genau ein Attribut.