6.3 Normalformen

Ist dies ein guter Schemaentwurf?

Stadt

SNr	SName	LCode	LFläche
7	Freiburg	D	357
9	Berlin	D	357
40	Moscow	RU	17075
43	St.Petersburg	RU	17075

- *Anomalie beim Einfügen*: Es können nur Länder aufgenommen werden, zu denen auch Städte existieren.
- *Anomalie beim Löschen*: Werden Städte gelöscht, können u.U. alle Informationen über gewisse Länder verloren gehen.
- *Anomalie beim Ändern*: Änderungen der Fläche eines Landes müssen bei mehreren Zeilen vorgenommen werden.

Zwei alternative Datenbankschemata.

Stadt

SNr	SName	LCode	LFläche
7	Freiburg	D	357
9	Berlin	D	357
40	Moscow	RU	17075
43	St.Petersburg	RU	17075

Stadt'

SNr	SName	LCode
7	Freiburg	D
9	Berlin	D
40	Moscow	RU
43	St.Petersburg	RU

Land'

LCode	LFläche
D	357
RU	17075

Gibt es hier Anomalien?

Sei $R = (V, \mathcal{F})$ ein Schema. Wir wollen eine Zerlegung $\rho = (X_1, \ldots, X_k)$ von R finden, die die folgenden Eigenschaften erfüllt:

- jedes $R_i = (X_i, \pi[X_i]\mathcal{F})$, $1 \leq i \leq k$ ist in einer gewünschten Normalform,
- ρ ist verlustfrei und (möglichst) auch abhängigkeitsbewahrend,
- k minimal.

Begriffe

- Sei X Schlüssel zu R und $X \subseteq Y \subseteq V$,
  dann nennen wir Y einen *Superschlüssel* von R.
- Gilt $A \in X$ für irgendeinen Schlüssel X von R,
  so bezeichnen wir A als *Schlüsselattribut (SA)* in R.
- Gilt $A \notin X$ für alle Schlüssel X,
  so bezeichnen wir A als *Nicht-Schlüsselattribut (NSA)*.

Hinweis: In der Literatur werden unterschiedliche Varianten von Normalformen betrachtet; wir beschränken uns auf die am häufigsten betrachteten.

3. Normalform

Ein Relationsschema $R = (V, \mathcal{F})$ ist in *3. Normalform* (3NF) genau dann, wenn jedes NSA $A \in V$ die folgende Bedingung erfüllt:

Wenn $X \to A \in \mathcal{F}$, $A \notin X$, dann ist X ein Superschlüssel.

Die Bedingung der 3NF verbietet nichttriviale funktionale Abhängigkeiten $X \to A$, in denen ein NSA A in der Weise von einem Schlüssel K transitiv funktional abhängt, dass $K \to X$ und $X \to A$, wobei $K \not\subseteq X$.

Welche Art von Redundanz wird so vermieden?
Redundanz der Art, dass mehrere Tupel mit denselben X-Werten existieren, so dass der Zusammenhang zu dem immer gleichen A-Wert redundant ist.

Welche funktionale Abhängigkeiten verletzen die 3NF?

Stadt

SNr	SName	LCode	LFläche
7	Freiburg	D	357
9	Berlin	D	357
40	Moscow	RU	17075
43	St.Petersburg	RU	17075

Kontinent

KName	LCode	KFläche	Prozent
Europe	D	3234	100
Europe	RU	3234	20
Asia	RU	44400	80

Was ist hier zu sagen?

Stadt'

SNr	SName	LCode
7	Freiburg	D
9	Berlin	D
40	Moscow	RU
43	St.Petersburg	RU

Land'

LCode	LFläche
D	357
RU	17075

Lage'

LCode	KName	Prozent
D	Europe	100
RU	Europe	20
RU	Asia	80

Kontinent'

KName	KFläche
Europe	3234
Asia	44400

Boyce-Codd-Normalform

Ein Relationsschema $R = (V, \mathcal{F})$ ist in *Boyce-Codd-Normalform* (BCNF) genau dann, wenn die folgende Bedingung erfüllt ist.

Wenn $X \to A \in \mathcal{F}$, $A \notin X$, dann ist X ein Superschlüssel.

Die BCNF verschärft die 3NF.

- Sei $R = (V, \mathcal{F})$
  mit $V = \{$ `Stadt, Adresse, PLZ` $\}$ und $\mathcal{F} = \{$ `Stadt Adresse` $\to$ `PLZ`, `PLZ` $\to$ `Stadt` $\}$
- R ist in 3NF aber nicht in BCNF.
- Sei $\rho = \{$`Adresse PLZ, PLZ Stadt`$\}$ eine Zerlegung.
  Dann erfüllt ρ die BCNF und ist verlustfrei, jedoch nicht abhängigkeitsbewahrend.

Bemerkung

Zu einem Relationsschema $R = (V, \mathcal{F})$ existiere genau einen Schlüssel.
R ist in BCNF genau dann, wenn R in 3NF.

6.3.1 Minimale Überdeckung: Basis für Normalisierungsalgorithmen

- Sei $\mathcal{F}$ eine Menge von funktionalen Abhängigkeiten.
- Wir suchen eine *"minimale" Überdeckung* von $\mathcal{F}$.

  $\mathcal{G}$ überdeckt $\mathcal{F}$, wenn $\mathcal{F} \equiv \mathcal{G}$, d.h. $\mathcal{F}^+ = \mathcal{G}^+$.

- *Strategie*:
  Bilde $\mathcal{G}$ durch Streichen von Attributen in den FAs von $\mathcal{F}$ oder Entfernen von FAs in $\mathcal{F}$ in einer Weise, die die Äquivalenz nicht zerstört.

Beispiel: Rechtsreduktion

- $\mathcal{F}_1 = \{B \to A, B \to C, A \to C\}$.

  Kann die FA $B \to C$ zu $B \to \emptyset$ reduziert werden, d.h. gestrichen werden?

  Sei $\mathcal{F}_1' = \{B \to A, A \to C\}$.

  Gilt $\mathcal{F}_1^+ = \mathcal{F}_1'^+$?

  Ja, denn
  (a) $\mathcal{F}_1^+ \supseteq \mathcal{F}_1'^+$ wegen $\mathcal{F}_1 \supseteq \mathcal{F}_1'$.
  (b) $\mathcal{F}_1^+ \subseteq \mathcal{F}_1'^+$ wegen $XPlus(B, C, \mathcal{F}_1')$.

Beispiel: Linksreduktion

- $\mathcal{F}_2 = \{AB \to C, B \to A\}$.

  Kann die FA $AB \to C$ zu $B \to C$ reduziert werden, d.h. $AB \to C$ durch $B \to C$ ersetzt werden?

  Sei $\mathcal{F}'_2 = \{B \to C, B \to A\}$.

  Gilt $\mathcal{F}_2^+ = \mathcal{F}_2'^+$?

  Ja, denn
  (a) $\mathcal{F}_2^+ \supseteq \mathcal{F}_2'^+$ wegen $XPlus(B, C, \mathcal{F}_2)$.
  (b) $\mathcal{F}_2^+ \subseteq \mathcal{F}_2'^+$ wegen (A2) und (A6) angewendet auf $B \to C$.

Links- und Rechtsreduktion

- Eine Menge $\mathcal{F}$ funktionaler Abhängigkeiten heißt *linksreduziert*, wenn sie die folgende Eigenschaft erfüllt:

  Wenn $X \to Y \in \mathcal{F}, Z \subset X$, dann
  $\mathcal{F}' = (\mathcal{F} \setminus \{X \to Y\}) \cup \{Z \to Y\}$ nicht äquivalent zu $\mathcal{F}$.

  [***Linksreduktion:*** ersetze $X \to Y$ in $\mathcal{F}$ durch $Z \to Y$.]

- $\mathcal{F}$ heißt *rechtsreduziert*, wenn sie die folgende Eigenschaft erfüllt:

  Wenn $X \to Y \in \mathcal{F}, Z \subset Y$, dann
  $\mathcal{F}' = (\mathcal{F} \setminus \{X \to Y\}) \cup \{X \to Z\}$ nicht äquivalent zu $\mathcal{F}$.

  [***Rechtsreduktion:*** ersetze $X \to Y$ in $\mathcal{F}$ durch $X \to Z$.]

Entscheidung mittels XPlus-Algorithmus

- Sei $X \to Y$ eine Abhängigkeit in $\mathcal{F}$. Betrachte $Z \to Y$, wobei $Z \subseteq X$.
  Wir führen die entsprechende *Linksreduktion* durch, wenn $\mathit{XPlus}(Z, Y, \mathcal{F}) = \texttt{true}$
- Sei $X \to Y$ eine Abhängigkeit in $\mathcal{F}$. Betrachte $X \to Z$, wobei $Z \subseteq Y$.
  Wir führen die entsprechende *Rechtsreduktion* durch, wenn $\mathit{XPlus}(X, Y, \mathcal{F}') = \texttt{true}$

Satz

Sei eine Menge funktionaler Abhängigkeiten $\mathcal{F}$ gegeben und sei $\mathcal{F}'$ aus $\mathcal{F}$ durch eine Links- oder Rechtsreduktion hervorgegangen.

Dann gilt: $\mathcal{F} \equiv \mathcal{F}'$.

Minimale Überdeckung

Eine Menge funktionaler Abhängigkeiten $\mathcal{F}^{min}$ ist eine *minimale Überdeckung* zu $\mathcal{F}$, wenn wir sie durch Anwendung der folgenden Schritte erzeugen können:

(1) Führe alle möglichen Linksreduktionen durch.

(2) Führe alle möglichen Rechtsreduktionen durch.

(3) Streiche alle trivialen funktionalen Abhängigkeiten der Form $X \to \emptyset$.

(4) Vereinige alle funktionalen Abhängigkeiten mit gleicher linker Seite $X \to Y_1, \ldots, X \to Y_n$ zu einer einzigen FA der Form $X \to Y_1 \ldots Y_n$.

6.3.2 Algorithmus zur Normalisierung

3NF-Synthese: verlustfrei und abhängigkeitsbewahrend

Sei $R = (V, \mathcal{F})$ ein Relationsschema.

1. Sei $\mathcal{F}^{min}$ eine minimale Überdeckung zu $\mathcal{F}$.
2. Betrachte jeweils maximale Klassen von funktionalen Abhängigkeiten aus $\mathcal{F}^{min}$ mit derselben linken Seite. Seien $\mathcal{C}_i = \{X_i \to A_{i1}, X_i \to A_{i2}, \ldots\}$, $i \geq 0$, die so gebildeten Klassen.[1]
3. Bilde zu jeder Klasse $\mathcal{C}_i$ ein Schema mit Format $V_{\mathcal{C}_i} = X_i \cup \{A_{i1}, A_{i2}, \ldots\}$.
4. Sofern keines der gebildeten Formate $V_{\mathcal{C}_i}$ einen Schlüssel für R enthält, berechne einen Schlüssel für R. Sei Y ein solcher Schlüssel. Bilde zu Y ein Schema mit Format $V_K = Y$.
5. $\rho = \{V_K, V_{\mathcal{C}_1}, V_{\mathcal{C}_2}, \ldots\}$ ist eine verlustfreie und abhängigkeitsbewahrende Zerlegung von R in 3NF.

[1] Der von uns betrachtete Algorithmus zur Berechnung von $\mathcal{F}^{min}$ hat diese Klassenbildung bereits vorgenommen, siehe (4) auf vorheriger Folie.