

Vorlesung 5

Normalisierung

- Normalformen definieren Qualitätskriterien
- Redundanz ist oft die Ursach von Schema-Problemen
- Normalisierung:
 - Jede Relation entspricht genau einer Objektmenge oder genau einer Relationship-Menge zwischen Objekten
 - Alle Informationen und Integritätsbedingungen (Schlüsselkandidaten und Fremdschlüsselelemente) sind abgebildet
 - Redundanz ist weitgehend eliminiert
 - Es treten keine Änderungsanomalien auf
 - ⇒ Beseitigung von Abhängigkeiten innerhalb der Relation
- Integritätsregeln, insbesondere FDs, helfen ein schlechtes Schema zu identifizieren
- die Lösung ist meistens Zerlegung der Relation

Zerlegung einer Relation

- Ist die Relation mit Blick auf FDS redundanzfrei?
 - Ja \Rightarrow keine Zerlegung notwendig
 - Nein \Rightarrow statt Zerlegungsprozedur

Zerlegungsprozedur

- Zerlegung der Relation R in R_1 und R_2 :
 - Ist dies informationserlustlos möglich?
 - Ja \Rightarrow OK
 - Nein \Rightarrow keine Zerlegung möglich
- Aufspaltung der zugehöriger FDS in $FD_{1,1}$ und $FD_{2,1}$, die jeweils R_1 und R_2 zugeordnet werden können
 - Ist dies abhängigkeitserhaltend möglich?
 - Ja \Rightarrow OK
 - Nein \Rightarrow OK, aber „unschön“, da Überprüfung der FDS nur nach der Rekonstruktion von R möglich ist (Effizienzverlust)

Normalformen

- Wenn eine Relation in einer Normalform ist, dann wissen wir, dass bestimmte Probleme nicht vorkommen können
- Die Normalformen basierend auf FDS:
$$BCNF \subseteq 3NF \subseteq 2NF \subseteq 1NF$$

- 1 Normalform (1NF)
 - > keine mehrwertigen Attribute
- 2 Normalform (2NF)
 - > keine Verbindung von Lachnerhalter in Relationen
- 3 Normalform (3NF)
 - > keine funktionale Abhängigkeiten von Nichtschlüsselattributen innerhalb von Relationen
- Boyce - Codd - Normalform (BCNF)
- Weitere Normalformen
 - > 4. Normalform (4NF)
 - > 5. Normalform (5NF)

1. Normalform

- Definition: Ein Relation ist in der ersten Normalform, wenn alle Attribute des Relation atomic sind.
Zusammengesetzte und mehrwertige Attribute sind nicht erlaubt
- Ist bei der genutzten Definition des relationalen Modells normalerweise automatisch eingehalten
- Atomic: String, Integer etc.

2. Normalform

Definition: Eine Relation R mit
zugehöriger FDS F ist in
Zweiter Normalform, genau dann
wenn es in 1NF ist und
jedes Nichtschlüsselattribut
 $A \in R$ voll funktional abhängig
ist von jedem Kandidatenschlüssel
der Relation

- Eine funktionale Abhängigkeit
 $X \rightarrow Y$ heißt voll, wenn es keine
echte Teilmenge $Z \subset X$ gibt, s.d $Z \rightarrow Y$
- Intuition: Ein Relationschema verletzt
2NF, wenn in der Relation
Informationen über mehr als ein
einziges Konzept modelliert werden
- Läßt bei Erfüllung der 2 NF können
immer noch Redundanzen im Schema
enthalten sein (durch transitive
Abhängigkeiten)

3. Normalform

- Definition: Eine Relation ist in 3NF, wenn sie in 2NF ist und kein Nichtschlüsselattribut von einem Schlüsselkandidaten transitiv abhängt
- Def 2: Eine Relation mit zugehörigen FDS F ist in der 3. Normalform wenn für alle Abhängigkeiten $A \rightarrow B$ aus F^+ gilt:
 - $B \subseteq A$ (FD ist trivial) oder
 - A enthält einer Schlüssel vor R (A ist ein Superschlüssel) oder
 - B ist Teil eines Schlüsselkandidaten (B ist prim)
- Bemerkung: 3NF beseitigt Abhängigkeiten vor Nicht-Schlüsselattributen

Boyce - Codd Normalform

- Definition: Eine Relation R mit zugehörigen FDS F ist in der Boyce - Codd Normalform, wenn für alle Abhängigkeiten $A \rightarrow B$ aus F^+ gilt:
 - $B \subseteq A$ (FD ist trivial) oder
 - A enthält einer Schlüssel vor R (A ist ein Superschlüssel)

→ Die BCNF hat also außer den trivialen nur noch funktionale Abhängigkeiten daher Determinante (linke Seite) ein Superschlüssel ist

Kanonische Überdeckung

$$\cdot F = \{ A B C D \rightarrow E, E \rightarrow D, A \rightarrow B, A C \rightarrow D \}$$

1. Linkssimplifikation

$$A B C D \rightarrow E$$

$$A B C D^+ = A B C D E$$

$$B C D^+ = B C D \quad \times$$

$$A C D^+ = A B C D E \quad \checkmark$$

$$A D^+ = A B D \quad \times$$

$$A C^+ = A B C D E \quad \checkmark$$

$$A C \rightarrow E$$

$$A C \rightarrow D$$

$$A^+ = A B \quad \times$$

$$C^+ = C \quad \times$$

$$F = \{ A C \rightarrow E, E \rightarrow D, A \rightarrow B, A C \rightarrow D \}$$

2. Rechts simplifikation

$$F = \{ A C \rightarrow E, E \rightarrow D, A \rightarrow B, A C \rightarrow \emptyset \}$$

$$3. \quad F = \{ A C \rightarrow E, E \rightarrow D, A \rightarrow B \}$$

4. —

$$\cdot F = \{ A \rightarrow B, B \rightarrow C, AB \rightarrow C \}$$

1. Linkssimplifikation

$$AB^T = ABC$$

$$A^T = ABC$$

$$A \rightarrow C$$

$$F = \{ A \rightarrow B, B \rightarrow C, A \rightarrow C \}$$

2. Rechts simplifikation

$$F = \{ A \rightarrow B, B \rightarrow \emptyset, A \rightarrow C \}$$

$$3. F = \{ A \rightarrow B, A \rightarrow C \}$$

$$4. F = \{ A \rightarrow BC \}$$