Vorlesung 4

Probleme (Anomalien)

→ Wegen schlechtes Schemadesign können folgende Probleme vorkommen:

- Einfüge-Anomalie
  - → Ursache: Vermischung zweier Entitätstypen
  - → Wirkung: Es können nicht Daten nur einer neuen Entitätstyp angelegt werden
- Lösch-Anomalie:
  - → Ursache: Vermischung zweier Entitätstypen
  - → Wirkung: Beim Löschen der Information zu einer Entität können Informationen zu einer andere Entität ungewollt verloren gehen
- Update-Anomalie:
  - → Ursache: Redundanz innerhalb der Relation
  - → Wirkung: Eine Aktualisierung kann zur Inkonsistenzen führen, wenn Änderung nicht in allen betroffenen Datensätze durchgeführt wird

Ziele des Datenbankentwurfs

- Vermeidung von Redundanzen und Anomalien:
  - -> Änderungen können bei Beachtung der Primärschlüssel und Fremdschlüssel keine Inkonsistenzen hervorrufen
  - -> Alle Informationen lassen sich unter Wahrung der Primärschlüssel- und Fremdschlüsselbedingungen einfügen
  - -> Informationen können einzeln wieder gelöscht werden, ohne die Primärschlüssel- oder Fremdschlussel-bedingungen zu verletzen
- Vermeidung von Informationsverlust
- Evtl. Einbeziehung von Effizienzüberlegung
- Vorgehen: Prozess der Zerlegung eines gegebenen Datenbank-Schemas in ein äquivalentes Schema ohne Redundanz und Anomalien
- Fragestellung bei der Entwurfstheorie:
  - -> wie kann die Güte eines Datenbankschemas beurteilt werden?
  - -> wie sieht ein gutes konzeptionelles Schema der Datenbank aus?
  - -> wie kann man ein schlechtes Schema in ein gutes Schema umwandeln?
- Erstellung „schöner" Relationenschemata:
  - -> Normalisierung = Überführung in eine redundanzarme Standard-Form
  - -> die Theorie der funktionalen Abhängigkeit

Funktionalle Abhängigkeit

→ Ein Attribut oder eine Kombination von Attributen bestimmt die Werte eines anderen Attributs oder Attributkombination

- Notation $A \to B$
  → A bestimmt B (funktional), oder A identifiziert oder impliziert B
- Eine Abhängigkeit $A \to B$ ist trivial, wenn $B \subseteq A$
- Eine funktionale Abhängigkeit $X \to Y$ heißt voll, wenn es keine echte Teilmenge $Z \subset X$ gibt, s.d $Z \to Y$
- Gibt es eine solche Teilmenge, dann heißt $X \to Y$ partielle Abhängigkeit

Berechnung funktionalen Abhängigkeiten

- Aus einer Menge F von FDs sind weitere FDs herleitbar

→ eine Hülle (closure) F^+ von F ist die Menge aller funktionalen Abhängigkeiten, die aus den funktionalen Abhängigkeiten in F ableitbar sind

- Armstrong Axiome:
  → Reflexivität: Sei $B \subseteq A$. Dann gilt stets $A \to B$
  → Verstärkung: Falls $A \to B$ gilt, dann gilt auch $A \cup C \to B \cup C$
  → Transitivität: Falls $A \to B$ und $B \to C$, dann gilt auch $A \to C$

- Armstrong Axiome sind
  → Korekt
  → Vollständig
- Erweiterung der Armstrong Axiome:
  - Vereinigungsregel: Falls $A \to B$ und $A \to C$, dann gilt $A \to B \cup C$
  - Dekompositionsregel: Falls $A \to B \cup C$ gilt, dann gilt auch $A \to B$ und $A \to C$
  - Pseudotransitivität: Falls $A \to B$ und $B \cup C \to D$, dann gilt auch $A \cup C \to D$

Superschlüssel / Oberschlüssel

- Was machen wir mit FDs?
  → Man kann einen Schlüssel für R berechnen
- Superschlüssel
  → In der Relation R ist $A \subseteq R$ eine Superschlüssel falls $A \to R$ gilt
  → d.h. A bestimmt alle andere Attributwerte der Relation R
  → Superschlüssel ist nicht notwendig minimal, $R \to R$ gilt immer

Kandidatenschlüssel

- Sei $K \subseteq R$
- K ist Kandidatenschlüssel falls:
  - K ist Superschlüssel
  - $\nexists K' \subset K$, s.d. K' Superschlüssel

→ Ein Attribut heißt prim in R, wenn es in einem Kandidatenschlüssel von R enthalten ist

Algorithmus Hülle (F, A)

- Input F, A
- Output : A^+

```
Erg := A;
While (Änderungen und Erg) Do
    Foreach FD (B->C) ∈ F Do
        If B ⊆ Erg Then Erg := Erg ∪ C;
Return A+ = Erg
```

Zerlegung (Dekomposition) eines Relationsschema:

- Um Anomalien zu beseitigen, wird das Schema RS einer Relation R in eine Vielzahl kleinerer Relationschemata $RS_1, \dots, RS_n$ zerlegt, so dass:
    - $RS_i \subseteq RS$, $1 \le i \le n$
    - $RS = \bigcup_i RS_i$
- Die Relation R wird in der Relationen $R_1, \dots, R_n$ zerlegt, wobei
    - $\rightarrow R_i = \pi_{RS_i}(R)$

Korrektheitskriterien

- → Verlustlosigkeit
- → Abhängigkeitserhaltung

Natürlichen Verbund (Natural Join)

- Das Kreuzprodukt wird gebildet, aus dem dann nur diejenigen Tupel selektiert werden, deren Attributwerte für gleichbenannte Attribute gleich sind
- Bezeichnung: $R \bowtie S$

Verlustlose Zerlegung

- Sei die Zerlegung der Relation R in $R_1, \dots, R_n$
- Es gilt immer: $r \subseteq \pi_{R_1}(r) \bowtie \dots \pi_{R_n}(r)$

- Die Zerlegung hat keinen Informationsverlust für jede gültige Ausprägung r in R gilt:

$$r = \pi_{R_1}(r) \bowtie \dots \bowtie \pi_{R_n}(r)$$

→ die in R erhaltene Information muss über den natürlichen Verbund der Relationen $R_1, \dots, R_n$ rekonstruierbar sein

Kriterien für die Verlustlosigkeit

- Theorem: Eine Zerlegung von R in R_1 und R_2 hat keinen Informationsverlust, falls einer der folgenden Bedingungen gilt:
  - $(R_1 \cap R_2 \rightarrow R_1) \in F_R^+$ oder
  - $(R_1 \cap R_2 \rightarrow R_2) \in F_R^+$, wobei F_R = Menge der FDs aus R
- Zerlegung Corollary: Wenn für R $\alpha \rightarrow \beta$, dann ist die Zerlegung $(R-\beta, \alpha \cup \beta)$ eine verlustlose Zerlegung

Verlustlose Zerlegung – Transitivität

- Theorem: Wenn $\{R_1, R_2\}$ eine verlustlose Zerlegung von R und $\{R_{11}, R_{12}\}$ eine verlustlose Zerlegung von R_1 sind, dann ist $\{R_{11}, R_{12}, R_2\}$ eine verlustlose Zerlegung von R

Abhängigkeitsbewahrung

- Idee: Alle FDs, die für die Relation R gelten, sollen lokal auf den Relationen $R_1, \ldots, R_n$ aus der Zerlegung überprüfbar sein
- Die Projektion von F auf α (F_α) ist die Menge von FDs aus F^+ die nur Attribute aus α enthalten:

$$F_\alpha = \{\beta \to \gamma \in F^+ \mid \beta\gamma \subseteq \alpha\}$$

Abhängigkeitsbewahrung:

- Berechnung von FD Projektionen:

Input: F, α

Output: F_α

Erg $= \emptyset$

For each $\beta \subseteq \alpha$ do

  $T = \beta^+$ (bzgl. F)

  Erg $=$ Erg $\cup \{\beta \to T \cap \alpha\}$

Return Erg

Hüllentreue Zerlegung

- Definition: Die Zerlegung der Relation R in Relationen $R_1, \ldots, R_n$ wird als hüllentreue Zerlegung bezeichnet falls:

$$F_R^+ = (F_{R_1} \cup \ldots \cup F_{R_n})^+$$

- $F_{R_1} \cup \ldots \cup F_{R_n} = F_R$

Zerlegung einer Relation

- Zerlegung notwendig
  - Ist die Relation mit Blick auf FDs redundanzfrei?
    - → Ja => keine Zerlegung notwendig
    - → Nein => Starte Zerlegungsprozedur
- Zerlegungsprozedur
  - → Zerlegung der Relation R in R_1 und R_2
  - Ist dies informationsverlustfrei (verlustlos) möglich:
    - → Ja => OK
    - → Nein => keine Zerlegung möglich
  - → Aufspaltung der zugehörigen FDs in FD1s und FD2s, die jeweils R_1 und R_2 zugeordnet werden könnten
  - Ist dies Abhängigkeitserhaltend möglich?
    - → Ja => OK
    - → Nein => OK, aber unschön, da Überprüfung der FDs nur nach der Rekonstruktion von R möglich ist (Effizienzverlust)