

Kapitel 6: Formaler Datenbankentwurf

- ▶ Die Schwierigkeiten der konzeptuellen Modellierung sind zu einem großen Teil dadurch begründet, dass sich die relevanten Strukturen einer Miniwelt erst in Diskussionen mit den Anwendern oder durch Analyse von Dokumenten erfassen lassen.
- ▶ Mit der Transformation eines konzeptuellen Schemas in ein relationales Schema, dem logischen Entwurf, ändert sich diese Situation.
- ▶ Die Konzepte des relationalen Datenmodells werden dahingehend erweitert, dass sich die Güte eines logischen Entwurfs formal überprüfen lässt.
- ▶ Grundlage hierfür sind sogenannte funktionale Abhängigkeiten.

6.1 Funktionale Abhängigkeiten

Definition

- ▶ Sei ein Relationsschema gegeben durch sein Format V und seien $X, Y \subseteq V$.
- ▶ Sei $\text{Rel}(V)$ die Menge aller möglichen Relationen über V .
- ▶ Sei $r \in \text{Rel}(V)$. r erfüllt eine funktionale Abhängigkeit (FA) $X \rightarrow Y$, wenn für alle $\mu, \nu \in r$ gilt:

$$\mu[X] = \nu[X] \Rightarrow \mu[Y] = \nu[Y].$$

Wir sagen auch die FA $X \rightarrow Y$ gilt in r .

	<i>A</i>	<i>B</i>	<i>C</i>
$r =$	a_1	b_1	c_1
	a_2	b_2	c_2
	a_2	b_2	c_3
	a_3	b_2	c_4

Welche FA werden durch r erfüllt?

FA	gilt in r
$A \rightarrow B$	1
$C \rightarrow A$	1
$C \rightarrow B$	1
$AC \rightarrow B$	1
$C \rightarrow AB$	1
$A \rightarrow C$	0
$B \rightarrow A$	0
$B \rightarrow C$	0
$AB \rightarrow C$	0
$A \rightarrow BC$	0
$B \rightarrow AC$	0
$A \rightarrow \emptyset$	1
$A \rightarrow A$	1
$AB \rightarrow A$	1
$ABC \rightarrow A$	1
\vdots	\vdots
$\emptyset \rightarrow A$	0
\vdots	\vdots
\vdots	\vdots

$\text{Sat}(V, \mathcal{F})$

- ▶ Sei \mathcal{F} eine Menge funktionaler Abhängigkeiten über V und $X, Y \subseteq V$.
- ▶ Sei dann $\text{Rel}(V)$ die Menge aller möglichen Relationen über V .
- ▶ Die Menge aller Relationen $r \in \text{Rel}(V)$, die alle funktionalen Abhängigkeiten in \mathcal{F} erfüllen, bezeichnen wir mit $\text{Sat}(V, \mathcal{F})$.

Membership-Test

- ▶ \mathcal{F} impliziert die funktionale Abhängigkeit $X \rightarrow Y$, $\mathcal{F} \models X \rightarrow Y$, wenn jede Relation $r \in \text{Sat}(V, \mathcal{F})$ auch $X \rightarrow Y$ erfüllt.
- ▶ Die Menge $\mathcal{F}^+ = \{X \rightarrow Y \mid \mathcal{F} \models X \rightarrow Y\}$ nennen wir die Hülle von \mathcal{F} .
- ▶ Der Test $X \rightarrow Y \in \mathcal{F}^+$ ist der Membership-Test.

Armstrong-Axiome

Sei $r \in \text{Sat}(V, \mathcal{F})$.

(A1) **Reflexivität:** Wenn $Y \subseteq X \subseteq V$, dann erfüllt r die FA $X \rightarrow Y$.

(A2) **Augmentation:**

Wenn $X \rightarrow Y \in \mathcal{F}, Z \subseteq V$, dann erfüllt r auch die FA $XZ \rightarrow YZ$.

(A3) **Transitivität:**

Wenn $X \rightarrow Y, Y \rightarrow Z \in \mathcal{F}$, dann erfüllt r auch die FA $X \rightarrow Z$.

(A1) erlaubt die Herleitung funktionaler Abhängigkeiten, ohne Bezug auf \mathcal{F} .
Wir bezeichnen solche FA als triviale funktionale Abhängigkeiten.

Korrektheit und Vollständigkeit

- ▶ Die Armstrong-Axiome sind korrekt in dem Sinn, dass die mit ihnen herleitbaren funktionalen Abhängigkeiten in der Tat Elemente der Hülle \mathcal{F}^+ sind.
- ▶ Die Armstrong-Axiome sind auch vollständig, d.h. jede funktionale Abhängigkeit in \mathcal{F}^+ kann auch mit ihnen hergeleitet werden.

Membership-Test (Variante 1)

Seien $X, Y \subseteq V$ und \mathcal{F} eine Menge funktionaler Abhängigkeiten.

Gilt $X \rightarrow Y \in \mathcal{F}^+$?

Starte mit \mathcal{F} und wende solange die Regeln (A1)–(A3) an bis entweder

- ▶ $X \rightarrow Y$ hergeleitet
oder
- ▶ \mathcal{F}^+ hergeleitet und $X \rightarrow Y \notin \mathcal{F}^+$

Ein solcher Algorithmus ist im Allgemeinen mindestens exponentiell in der Anzahl der Attribute in V , da für (A2) alle Teilmengen von V betrachtet werden müssen.

weitere Axiome

Seien $X, Y, Z, W \subseteq V$ und $A \in V$.

(A4) Vereinigung:

Wenn $X \rightarrow Y, X \rightarrow Z \in \mathcal{F}$, dann erfüllt r auch die FA $X \rightarrow YZ$.

(A5) Pseudotransitivität:

Wenn $X \rightarrow Y, WY \rightarrow Z \in \mathcal{F}$, dann erfüllt r auch die FA $XW \rightarrow Z$.

(A6) Dekomposition:

Wenn $X \rightarrow Y \in \mathcal{F}, Z \subseteq Y$, dann erfüllt r auch die FA $X \rightarrow Z$.

(A7) Reflexivität:

Wenn $X \subseteq V$, dann erfüllt r auch die FA $X \rightarrow X$.

(A8) Akkumulation:

Wenn $X \rightarrow YZ, Z \rightarrow AW \in \mathcal{F}$, dann erfüllt r auch die FA $X \rightarrow YZA$.

Die Axiomensysteme $\{(A1), (A2), (A3)\}$ und $\{(A6), (A7), (A8)\}$ sind zueinander äquivalent.

Hierzu müssen wir zeigen, dass jedes Axiom der einen Menge durch die Axiome der anderen Menge simuliert werden kann.

Beispiel: (A6) kann durch (A1) und (A3) simuliert werden

Seien $X, Y, Z \subseteq Y$.

Zu zeigen: $X \rightarrow Y, Z \subseteq Y \Rightarrow X \rightarrow Z$.

$$\begin{array}{rcl} Z \subseteq Y & \stackrel{A1}{\Rightarrow} & Y \rightarrow Z \\ X \rightarrow Y, Y \rightarrow Z & \stackrel{A3}{\Rightarrow} & X \rightarrow Z \end{array}$$

Beispiel: (A8) kann durch (A1), (A2) und (A3) simuliert werden

Seien $X, Y, Z, W \subseteq V$ und $A \in V$.

Zu zeigen: $X \rightarrow YZ, Z \rightarrow AW \in \mathcal{F} \Rightarrow X \rightarrow YZA$.

$$\begin{array}{rcl} Z \rightarrow AW & \stackrel{A2}{\Rightarrow} & YZ \rightarrow YZAW \\ X \rightarrow YZ, YZ \rightarrow YZAW & \stackrel{A3}{\Rightarrow} & X \rightarrow YZAW \\ X \rightarrow YZAW & \stackrel{A1, A3}{\Rightarrow} & X \rightarrow YZA \end{array}$$

(Attribut-)Hülle X^+ von X (bzgl. \mathcal{F})

Sei $X \subseteq V$.

Die Hülle von X , bezeichnet als X^+ , ist definiert als

$$X^+ = \{A \mid A \in V \text{ und } X \rightarrow A \in \mathcal{F}^+\}.$$

Membership-Test (Variante 2)

Seien $X, Y \subseteq V$ und \mathcal{F} eine Menge funktionaler Abhängigkeiten.

Gilt $X \rightarrow Y \in \mathcal{F}^+$?

Berechne zunächst X^+ mittels (A6) - (A8) und teste anschließend, ob $Y \subseteq X^+$.

XPlus-Algorithmus

```
XPlus( $X, Y, \mathcal{F}$ ) boolean {  
  result :=  $X$   
  WHILE (changes to result) DO  
    FOREACH  $X' \rightarrow Y' \in \mathcal{F}$  DO  
      IF ( $X' \subseteq \text{result}$ ) THEN result := result  $\cup$   $Y'$   
    END  
  END  
  IF ( $Y \subseteq \text{result}$ ) RETURN true ELSE false  
}
```

Der XPlus-Algorithmus hat eine Laufzeit, die polynomiell in der Darstellung von \mathcal{F} ist.

Beispiel XPlus-Algorithmus

Sei $V = \{A, B, C, D, E, F, G, H, I\}$ und
 $\mathcal{F} = \{AB \rightarrow E, BE \rightarrow I, E \rightarrow G, GI \rightarrow H\}$.

Es soll getestet werden, ob $AB \rightarrow GH \in \mathcal{F}^+$.

Axiom	Anwendung	result
(A7)	$AB \rightarrow AB$	$\{A, B\}$
(A8)	$AB \rightarrow ABE$	$\{A, B, E\}$
(A8)	$AB \rightarrow ABEI$	$\{A, B, E, I\}$
(A8)	$AB \rightarrow ABEIG$	$\{A, B, E, I, G\}$
(A8)	$AB \rightarrow ABEIGH$	$\{A, B, E, I, G, H\}$
(A6)	$AB \rightarrow GH$	

Schlüssel - jetzt formal definiert:

Sei $V = \{A_1, \dots, A_n\}$. $X \subseteq V$ heißt Schlüssel für V (bzgl. \mathcal{F}), wenn

- (1) $X \rightarrow A_1 \dots A_n \in \mathcal{F}^+$,
- (2) $Y \subset X \Rightarrow Y \rightarrow A_1 \dots A_n \notin \mathcal{F}^+$.

Mit dem XPlus-Algorithmus können wir zu gegebenen V, \mathcal{F} einen Schlüssel berechnen.

- (1) Beginne mit $X := V$.
- (2) Betrachte die einzelnen $A \in V$ in einer beliebigen Reihenfolge:
Falls $(X \setminus \{A\})^+ = V$, dann $X := X \setminus \{A\}$
- (3) Lässt sich dieses Vorgehen nicht weiter fortsetzen, dann ist X ein Schlüssel.

Bemerkung

- ▶ Jedes $A \in V$ wird nur einmal betrachtet
 \Rightarrow der XPlus-Algorithmus wird folglich n -mal aufgerufen.
- ▶ Sofern mehrere Schlüssel existieren, z.B.
 $V = \{\text{Stadt}, \text{Adresse}, \text{PLZ}\}$, $\mathcal{F} = \{\text{Stadt} \rightarrow \text{Adresse}, \text{Adresse} \rightarrow \text{PLZ}, \text{PLZ} \rightarrow \text{Stadt}\}$,
wird nur einer dieser berechnet.

(Stadt, Adresse) und (Adresse, PLZ) sind Schlüssel zu R .
- ▶ ~~Der Test, ob eine beliebige gegebene Attributmenge ein Schlüssel ist, ist exponentiell (NP-vollständig).~~

Der Test, ob ein beliebiges gegebenes Attribut Teil eines Schlüssels ist, ist exponentiell (NP-vollständig).

6.2 Verlustfreie und abhängigkeitsbewahrende Zerlegungen

Zwei alternative Datenbankschemata:

Stadt			
<u>SNr</u>	SName	LCode	LFläche
7	Freiburg	D	357
9	Berlin	D	357
40	Moscow	RU	17075
43	St.Petersburg	RU	17075

Stadt'		
<u>SNr</u>	SName	LCode
7	Freiburg	D
9	Berlin	D
40	Moscow	RU
43	St.Petersburg	RU

Land'	
<u>LCode</u>	LFläche
D	357
RU	17075

Stadt' und Land' bilden eine Zerlegung von Stadt. Mit welchen Eigenschaften?

Eigenschaften von Zerlegungen

- ▶ Sei ein Relationsschema R gegeben durch eine Attributmenge V und eine Menge funktionaler Abhängigkeiten \mathcal{F} .

Sei $\rho = \{X_1, \dots, X_k\}$ eine Zerlegung von V , d.h. $X_i \subseteq V$, $\cup_{1 \leq i \leq k} X_i = V$.

- ▶ ρ ist verlustfrei, wenn jedes $r \in \text{Sat}(V, \mathcal{F})$ mittels \bowtie aus den einzelnen Relationen $\pi[X_i]r$ exakt rekonstruierbar ist.
- ▶ ρ abhängigkeitsbewahrend, wenn die funktionalen Abhängigkeiten in \mathcal{F} auch über den Schemata der Zerlegung ρ ausgedrückt werden können.

6.2.1 Verlustfreiheit

Definition: *Verlustfreie Zerlegung*

Sei $\rho = \{X_1, \dots, X_k\}$ eine Zerlegung von V , d.h. $X_i \subseteq V$, $\cup_{1 \leq i \leq k} X_i = V$.
 ρ heißt verlustfrei, wenn für jede Relation $r \in \text{Sat}(V, \mathcal{F})$ gilt:

$$r = \pi[X_1]r \bowtie \dots \bowtie \pi[X_k]r$$

Beispiel

- ▶ Sei $V = \{A, B, C\}$ und $\mathcal{F} = \{A \rightarrow B, A \rightarrow C\}$.
- ▶ Sei beispielsweise $r \in \text{Sat}(V, \mathcal{F})$ wie folgt:

$$r = \begin{array}{ccc} A & B & C \\ \hline a_1 & b_1 & c_1 \\ a_2 & b_1 & c_2 \end{array}$$

- ▶ Seien $\rho_1 = \{AB, BC\}$ und $\rho_2 = \{AB, AC\}$.
- ▶ $r \subset \pi[AB]r \bowtie \pi[BC]r$,
 ρ_1 ist nicht verlustfrei
- ▶ $r = \pi[AB]r \bowtie \pi[AC]r$,
 ρ_2 ist verlustfrei
(denn die Bedingung ist für beliebiges $r \in \text{Sat}(V, \mathcal{F})$ ebenfalls erfüllt)

Satz

Sei V eine Attributmenge mit einer Menge \mathcal{F} funktionaler Abhängigkeiten.
Sei $\rho = (X_1, X_2)$ eine Zerlegung von V .

ρ ist verlustfrei genau dann, wenn

$$(X_1 \cap X_2) \rightarrow (X_1 \setminus X_2) \in \mathcal{F}^+, \text{ oder } (X_1 \cap X_2) \rightarrow (X_2 \setminus X_1) \in \mathcal{F}^+.$$

Korollar

Sei $R = (V, \mathcal{F})$ ein Relationsschema und sei $X \rightarrow Y \in \mathcal{F}$, wobei $X \cap Y = \emptyset$.

Dann ist die Zerlegung $\rho = (V \setminus Y, XY)$ verlustfrei.

Beweis: $(V \setminus Y) \cap XY = X$; $XY \setminus (V \setminus Y) = Y$.

Beispiel

- ▶ Sei $V = \{A, B, C\}$ und $\mathcal{F} = \{A \rightarrow C\}$. $\rho = \{AB, AC\}$ ist verlustfrei.
- ▶ Sei $V = \{A, B, C\}$ und $\mathcal{F} = \{A \rightarrow B\}$. $\rho = \{AB, AC\}$ ist verlustfrei.

6.2.2 Abhängigkeitsbewahrung

Wenn die funktionalen Abhängigkeiten in \mathcal{F} auch über den Schemata einer Zerlegung ρ ausgedrückt werden können, dann nennen wir ρ abhängigkeitsbewahrend.

Beispiel

Sei $V = \{A, B, C, D\}$ und $\rho = \{AB, BC\}$.

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

Ist ρ abhängigkeitsbewahrend bzgl. \mathcal{F}_1 ?

- Betrachte $\mathcal{F}_2 = \{A \rightarrow B, B \rightarrow C, C \rightarrow B, B \rightarrow A\}$

Ist ρ abhängigkeitsbewahrend bzgl. \mathcal{F}_2 ?

Seien $\mathcal{F}, \mathcal{F}'$ Mengen funktionaler Abhängigkeiten. \mathcal{F} und \mathcal{F}' heißen äquivalent genau dann, wenn ihre Hüllen gleich sind, d.h. $\mathcal{F} \equiv \mathcal{F}'$ gdw. $\mathcal{F}^+ = \mathcal{F}'^+$.

- Es gilt $\mathcal{F}_1^+ = \mathcal{F}_2^+$.
- Es gilt also $\mathcal{F}_1 \equiv \mathcal{F}_2$ und somit
 ρ abhängigkeitsbewahrend sowohl bzgl. \mathcal{F}_1 als auch bzgl. \mathcal{F}_2 .

Definition: *Abhängigkeitsbewahrende Zerlegung*

- ▶ Sei $R = (V, \mathcal{F})$ gegeben. Sei weiter $Z \subseteq V$.
- ▶ Sei die Projektion von \mathcal{F} auf Z definiert zu

$$\pi[Z]\mathcal{F} = \{X \rightarrow Y \in \mathcal{F}^+ \mid XY \subseteq Z\}.$$

- ▶ Eine Zerlegung $\rho = \{X_1, \dots, X_k\}$ von V heißt abhängigkeitsbewahrend bzgl. \mathcal{F} , wenn

$$\bigcup_{i=1}^k \pi[X_i]\mathcal{F} \equiv \mathcal{F}.$$

zum vorangehenden Beispiel

Sei $V = \{A, B, C, D\}$, $\rho = \{AB, BC\}$ und $\mathcal{F} = \{A \rightarrow B, B \rightarrow C, C \rightarrow A\}$.

Ist ρ abhängigkeitsbewahrend bzgl. \mathcal{F} ?

Ja, denn

► $\{A \rightarrow B, B \rightarrow A\} \subseteq \pi[AB]\mathcal{F}$,

► $\{B \rightarrow C, C \rightarrow B\} \subseteq \pi[BC]\mathcal{F}$

und

► $\{A \rightarrow B, B \rightarrow A\} \cup \{B \rightarrow C, C \rightarrow B\} \equiv \mathcal{F}$.

Beobachtung: Nicht jede verlustfreie Zerlegung ist abhängigkeitsbewahrend!

- ▶ $R = (V, \mathcal{F})$, wobei $V = \{\text{Stadt, Adresse, PLZ}\}$,
- ▶ $\mathcal{F} = \{\text{Stadt Adresse} \rightarrow \text{PLZ}, \text{PLZ} \rightarrow \text{Stadt}\}$.
- ▶ $\rho = \{X_1, X_2\}$: $X_1 = \{\text{Adresse, PLZ}\}$ und $X_2 = \{\text{Stadt, PLZ}\}$.
- ▶ ρ ist verlustfrei, da $(X_1 \cap X_2) \rightarrow (X_2 \setminus X_1) \in \mathcal{F}$.
- ▶ ρ ist nicht abhängigkeitsbewahrend.