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 $\mathrm{Rel}(V)$ die Menge aller möglichen Relationen über V.
- Sei $r \in \mathrm{Rel}(V)$. r erfüllt eine *funktionale Abhängigkeit (FA)* $X \to Y$, wenn für alle $\mu, \nu \in r$ gilt:

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

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

$$r = \begin{array}{ccc} A & B & C \\ \hline a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_2 & b_2 & c_3 \\ a_3 & b_2 & c_4 \end{array}$$

Welche FA werden durch r erfüllt?

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

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

- Sei $\mathcal{F}$ eine Menge funktionaler Abhängigkeiten über V und $X, Y \subseteq V$.
- Sei dann 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 Sat($V, \mathcal{F}$).

Membership-Test

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

Armstrong-Axiome

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

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

(A2) **Augmentation**:
Wenn $X \to Y \in \mathcal{F}, Z \subseteq V$, dann erfüllt r auch die FA $XZ \to YZ$.

(A3) **Transitivität**:
Wenn $X \to Y, Y \to Z \in \mathcal{F}$, dann erfüllt r auch die FA $X \to Z$.

(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 $X \to Y \in \mathcal{F}^+$?

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

- $X \to Y$ hergeleitet

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

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

(A5) **Pseudotransitivität**:
Wenn $X \to Y, WY \to Z \in \mathcal{F}$, dann erfüllt r auch die FA $XW \to Z$.

(A6) **Dekomposition**:
Wenn $X \to Y \in \mathcal{F}, Z \subseteq Y$, dann erfüllt r auch die FA $X \to Z$.

(A7) **Reflexivität**:
Wenn $X \subseteq V$, dann erfüllt r auch die FA $X \to X$.

(A8) **Akkumulation**:
Wenn $X \to YZ, Z \to AW \in \mathcal{F}$, dann erfüllt r auch die FA $X \to YZA$.

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: $X \to Y, Z \subseteq Y \Rightarrow X \to Z$.

$$\begin{aligned} Z \subseteq Y &\quad \stackrel{A1}{\Rightarrow} \quad Y \to Z \\ X \to Y, Y \to Z &\quad \stackrel{A3}{\Rightarrow} \quad X \to Z \end{aligned}$$

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

Seien $X, Y, Z, W \subseteq V$ und $A \in V$.
Zu zeigen: $X \to YZ, Z \to AW \in \mathcal{F} \Rightarrow X \to YZA$.

$$\begin{aligned} Z \to AW &\quad \stackrel{A2}{\Rightarrow} \quad YZ \to YZAW \\ X \to YZ, YZ \to YZAW &\quad \stackrel{A3}{\Rightarrow} \quad X \to YZAW \\ X \to YZAW &\quad \stackrel{A1,A3}{\Rightarrow} \quad X \to YZA \end{aligned}$$

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

Sei $X \subseteq V$.
Die Hülle von X, bezeichnet als X^+, ist definiert als

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

Membership-Test (Variante 2)

Seien $X, Y \subseteq V$ und $\mathcal{F}$ eine Menge funktionaler Abhängigkeiten.
Gilt $X \to Y \in \mathcal{F}^+$?

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

XPlus-Algorithmus

```
XPlus(X,Y,F) boolean {
  result := X
  WHILE (changes to result) DO
    FOREACH X' → Y' ∈ F DO
       IF (X' ⊆ result) THEN result := result ∪ Y'
    END
  END
  IF (Y ⊆ result) RETURN true ELSE false
}
```

Beispiel XPlus-Algorithmus

Sei $V = \{A, B, C, D, E\}$ und $\mathcal{F} = \{A \to B, B \to C, C \to D, D \to E\}$.

Es soll getestet werden, ob $A \to CE \in \mathcal{F}^+$.

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
$\mathcal{F} = \{AB \to E, BE \to I, E \to G, GI \to H\}$.

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

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

Schlüssel - jetzt formal definiert:

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

(1) $X \to A_1 \dots A_n \in \mathcal{F}^+$,

(2) $Y \subset X \Rightarrow Y \to A_1 \dots A_n \notin \mathcal{F}^+$.

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.
  $V = \{\texttt{Stadt, Adresse, PLZ}\}$, $\mathcal{F} = \{\texttt{Stadt Adresse} \rightarrow \texttt{PLZ, PLZ} \rightarrow \texttt{Stadt}\}$,
  wird nur einer dieser berechnet.
- Der Test, ob eine beliebige gegebene Attributmenge ein Schlüssel 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 $\rho = \{X_1, \ldots, X_k\}$ eine *Zerlegung* von V, d.h. $X_i \subseteq V$, $\cup_{1 \leq i \leq k} X_i = V$.
- ρ ***ist verlustfrei, wenn:***
  Sei $r \in \mathrm{Sat}(V, \mathcal{F})$ und seien $r_i = \pi[X_i]r$, $1 \leq i \leq k$ die Projektionen von r auf die einzelnen Elemente der Zerlegung.

  r ist mittels $\bowtie$ aus den einzelnen Relationen der Zerlegung ρ exakt rekonstruierbar.
- ρ ***abhängigkeitsbewahrend, wenn:***
  Die funktionalen Abhängigkeiten in $\mathcal{F}$ können auch über den Schemata der Zerlegung ρ ausgedrückt werden.

6.2.1 Verlustfreiheit

Definition: *Verlustfreie Zerlegung*

Sei $\rho = \{X_1, \ldots, X_k\}$ 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 $r \in \mathrm{Sat}(V, \mathcal{F})$ gilt:

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

Beispiel

- Sei $V = \{A, B, C\}$ und $\mathcal{F} = \{A \to B, A \to C\})$.
- Sei beispielsweise $r \in Sat(V, \mathcal{F})$ 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 $r \in Sat(V, \mathcal{F})$ 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

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

Korollar

Sei $R = (V, \mathcal{F})$ ein Relationsschema und sei $X \to Y \in \mathcal{F}$, 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 $\mathcal{F} = \{A \to B, A \to C\})$. $\rho = \{AB, AC\}$ ist verlustfrei.
- Sei $V = \{A, B, C\}$ und $\mathcal{F} = \{A \to B\})$. $\rho = \{AB, AC\}$ 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 $\mathcal{F}_1 = \{A \to B, B \to C, C \to A\}$.
  Ist ρ abhängigkeitsbewahrend bzgl. $\mathcal{F}_1$?
- Betrachte $\mathcal{F}_2 = \{A \to B, B \to C, C \to B, B \to A\}$
  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 $\rho = \{X_1, \ldots, X_k\}$ 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 $\mathcal{F} = \{A \to B, B \to C, C \to A\}$.
Ist ρ abhängigkeitsbewahrend bzgl. $\mathcal{F}$?

Ja, denn

- $\{A \to B, B \to A\} \subseteq \pi[AB]\mathcal{F}$,
- $\{B \to C, C \to B\} \subseteq \pi[BC]\mathcal{F}$

  und
- $\{A \to B, B \to A\} \cup \{B \to C, C \to B\} \equiv \mathcal{F}$.

Beobachtung: Nicht jede verlustfreie Zerlegung ist abhängigkeitsbewahrend!

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

(Stadt, Adresse) und (Adresse, PLZ) sind Schlüssel zu R.