Kapitel 7: 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.
- Das relationale Datenmodell wird dahingehend erweitert, dass sich die Güte eines logischen Entwurfs formal überprüfen läßt.

7.1 Motivation

Relationen mit Anomalien

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

Relationen ohne Anomalien

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

Lage'

<u>LCode</u>	<u>KName</u>	Prozent
D	Europe	100
RU	Europe	20
RU	Asia	80

Kontinent'

<u>KName</u>	KFläche
Europe	3234
Asia	44400

7.2 Funktionale Abhängigkeiten

7.2.1 Definition

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

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

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

7.2.2 Membership-Test

- $\mathcal{F}$ *impliziert* die funktionale Abhängigkeit $X \to Y$, $\mathcal{F} \models X \to Y$, wenn jede Relation $r \in \mathrm{Sat}(V, \mathcal{F})$ 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*.

formale Definition: Schlüssel

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

- $X \rightarrow A_1 \ldots A_n \in \mathcal{F}^+$,
- $Y \subset X \Rightarrow Y \rightarrow A_1 \ldots A_n \notin \mathcal{F}^+$.

Armstrong-Axiome

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

(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 $\mathcal{F}$: *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.
- Sind die Armstrong-Axiome auch *vollständig*, d.h., kann jede funktionale Abhängigkeit in $\mathcal{F}^+$ auch mit ihnen hergeleitet werden?
  - *(Attribut-)Hülle* X^+ von X (bzgl. $\mathcal{F}$):

$$X^+ = \{A \quad | \; A \in V \text{ und } X \to A \text{ kann mittels (A1) – (A3) hergeleitet werden}\}.$$

  - Zum Nachweis der Vollständigkeit zeigen wir: Wenn $X \to Y$ nicht herleitbar mittels (A1)–(A3), dann $X \to Y \notin \mathcal{F}^+$, d.h. $\exists r$, r erfüllt $\mathcal{F}$, jedoch nicht $X \to Y$.

Membership-Test Variante 1:

Starte mit $\mathcal{F}$ und wende solange die Regeln (A1)–(A3) an, bis entweder $X \rightarrow Y$ hergeleitet, oder $\mathcal{F}^+$ hergeleitet und $X \rightarrow Y \notin \mathcal{F}^+$.

Welche Zeitkomplexität hat dieser Algorithmus?

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.

Beweise!

Membership-Test Variante 2:

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
    FOR each X' → Y' ∈ 𝓕 DO
       IF (X' ⊆ result) THEN result := result ∪ Y';
    end.
  IF (Y ⊆ result) RETURN true ELSE false;
}
```

Der XPlus-Algorithmus hat eine Laufzeit, die quadratisch in der Anzahl der funktionalen Abhängigkeiten und der Anzahl Attribute ist.

Warum?

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)	$AB \rightarrow AB$	$\{A, B\}$
...	...	...

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

Wie?

7.2.3 Minimale Überdeckung

Äquivalenz zweier Mengen von funktionalen Abhängigkeiten

- Seien $\mathcal{F}, \mathcal{G}$ Mengen von funktionalen Abhängigkeiten.
- $\mathcal{F}, \mathcal{G}$ nennen wir *äquivalent*, $\mathcal{F} \equiv \mathcal{G}$, wenn $\mathcal{F}^+ = \mathcal{G}^+$.
- Wir suchen eine *minimale Überdeckung* von $\mathcal{F}$.

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 $Z \to Y$.

- Sie heißt *rechtsreduziert*, 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 $\mathcal{F}$ und sei $Z \to Y$, wobei $Z \subset X$.

  Wir führen die entsprechende Linksreduktion durch, wenn $\mathit{XPlus}(Z, Y, \mathcal{F})$ das Ergebnis `true` liefert.
- Sei $X \to Y$ eine Abhängigkeit in $\mathcal{F}$ und sei $X \to Z$, wobei $Z \subset Y$.

  Wir führen die entsprechende Rechtsreduktion durch, wenn $\mathit{XPlus}(X, Y, \mathcal{F}')$ das Ergebnis `true` liefert.

Satz

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

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

Beispiel

- $\mathcal{F}_1 = \{A \to B, B \to A, B \to C, A \to C, C \to A\}$.
  Rechtsreduktionen?
- $\mathcal{F}_2 = \{AB \to C, A \to B, B \to A\}$.
  Linksreduktionen?

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:

- Führe alle möglichen Linksreduktionen durch.
- Führe alle möglichen Rechtsreduktionen durch.
- Streiche alle trivialen funktionalen Abhängigkeiten der Form $X \rightarrow \emptyset$.
- Vereinige alle funktionalen Abhängigkeiten mit gleicher linker Seite $X \rightarrow Y_1, \ldots, X \rightarrow Y_n$ zu einer einzigen FA der Form $X \rightarrow Y_1 \ldots Y_n$.

- Zu einer gegebenen Menge $\mathcal{F}$ funktionaler Abhängigkeiten können wir in polynomieller Zeit eine minimale Überdeckung $\mathcal{F}^{min}$ bestimmen.

  Wie?

- $\mathcal{F}^{min}$ ist jedoch im Allgemeinen nicht eindeutig.

  Warum?

7.3 Verlustfreie und abhängigkeitsbewahrende Zerlegungen

- Sei ein Relationsschema gegeben durch eine Attributmenge V und eine Menge funktionaler Abhängigkeiten $\mathcal{F}$.
- Sei $\rho = \{X_1, \ldots, X_k\}$ eine *Zerlegung* von V.
- 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.

Verlustfreiheit: Jede Relation zu dem Ausgangsschema bleibt mittels $\bowtie$ aus den einzelnen Relationen der Zerlegung ρ exakt rekonstruierbar.

Abhängigkeitsbewahrung: Die funktionalen Abhängigkeiten in $\mathcal{F}$ können auch über den Schemata der Zerlegung ρ ausgedrückt werden.

7.3.1 Verlustfreiheit

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 $r \in Sat(V, \mathcal{F})$ wie folgt:

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

- Sei $\rho_1 = \{AB, BC\}$ und $\rho_2 = \{AB, AC\}$.
- $r \qquad \pi[AB]r \bowtie \pi[BC]r$,
- $r \qquad \pi[AB]r \bowtie \pi[AC]r$.

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

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

7.3.2 Abhängigkeitsbewahrung

Beispiel

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

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

  Ist ρ abhängigkeitsbewahrend bzgl. $\mathcal{F}$?
- Betrachte die zu $\mathcal{F}$ äquivalente Menge $\mathcal{F}' = \{A \to B, B \to C, C \to B, B \to A\}$.

  Ist ρ abhängigkeitsbewahrend bzgl. $\mathcal{F}'$?

Definition

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

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

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

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

Nicht jede verlustfreie Zerlegung ist abhängigkeitsbewahrend!

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

Gib die Schlüssel zu R an!

7.4 Normalformen

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 der 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 *Superschlüssel* von R.
- Gilt $A \in X$ für irgendeinen Schlüssel X von R, so heißt A *Schlüsselattribut (SA)* in R;
- gilt $A \notin X$ für jeden Schlüssel X, so heißt A *Nicht-Schlüsselattribut (NSA)*.

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

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

Welche Art von Redundanz wird so vermieden?

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 $R = (V, \mathcal{F})$, wobei $V = \{$ `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, Stadt PLZ`$\}$ eine Zerlegung, dann erfüllt ρ die BCNF und ist verlustfrei, jedoch nicht abhängigkeitsbewahrend.

Satz

Zu einem Relationsschema $R = (V, \mathcal{F})$ existiere genau einen Schlüssel.

R ist in BCNF genau dann, wenn R in 3NF.

Satz

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

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

7.5 Algorithmen zur Normalisierung

BCNF-Analyse: verlustfrei und nicht abhängigkeitsbewahrend

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

1. Sei $X \subset V$, $A \in V$ und $X \rightarrow A \in \mathcal{F}$ eine FA, die die BCNF-Bedingung verletzt. Sei weiter $V' = V \setminus \{A\}$.

   Zerlege R in

   $$R_1 = (V', \pi[V']\mathcal{F}), \quad R_2 = (XA, \pi[XA]\mathcal{F}).$$

2. Teste die BCNF-Bedingung bzgl. R_1 und R_2 und wende den Algorithmus gegebenenfalls rekursiv an.

3NF-Analyse: verlustfrei und abhängigkeitsbewahrend

Sei $R = (V, \mathcal{F})$ ein Relationsschema und sei $\rho = (X_1, \ldots, X_k)$ eine Zerlegung von V, so dass die entsprechenden Schemata $R_1 = (X_1, \pi[X_1]\mathcal{F})$, ..., $R_k = (X_k, \pi[X_k]\mathcal{F})$ in BCNF.

1. Sei $\mathcal{F}^{min}$ eine minimale Überdeckung zu $\mathcal{F}$.
2. Identifiziere die Menge $\mathcal{F}' \subseteq \mathcal{F}^{min}$ derjenigen funktionalen Abhängigkeiten, für die die Abhängigkeitsbewahrung verletzt ist.
3. Für jede solche FA der Form $X \rightarrow A$ erweitere ρ um XA; die entsprechenden Schemata haben die Form $(XA, \pi[XA]\mathcal{F})$.

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 $\mathcal{C}_i = \{X_i \to A_{i1}, X_i \to A_{i2}, \ldots\}$ die so gebildeten Klassen, $i \geq 0$.
3. Bilde zu jeder Klasse $\mathcal{C}_i$, $i \geq 0$, ein Schema mit Format $V_{\mathcal{C}_i} = X_i \cup \{A_{i1}, A_{i2}, \ldots\}$.
4. Sofern keines der gebildeten Formaten $V_{\mathcal{C}_i}$ 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.