

Thomas Studer
Relationale Datenbanken:
Von den theoretischen Grundlagen
zu Anwendungen mit PostgreSQL
2. Auflage
Springer Vieweg, 2019
ISBN 978-3-662-58975-5

Dieser Foliensatz darf frei verwendet werden unter der Bedingung, dass diese Titelfolie nicht entfernt wird.

Datenbanken



Berechnung von Normalformen

Thomas Studer

Institut für Informatik
Universität Bern

Armstrong-Kalkül

Gegeben sei eine Menge F von funktionalen Abhängigkeiten über U . Dann wird $F \vdash X \rightarrow Y$ für $X, Y \subseteq U$ induktiv definiert durch:

1. **Elemente von F .** Ist $X \rightarrow Y$ ein Element von F , so gilt $F \vdash X \rightarrow Y$.
2. **Reflexivität.** Ist Y eine Teilmenge von X , so gilt $F \vdash X \rightarrow Y$.
3. **Augmentation.** Aus $F \vdash X \rightarrow Y$ und $Z \subseteq U$ folgt $F \vdash XZ \rightarrow YZ$.
4. **Transitivität.** Aus $F \vdash X \rightarrow Y$ und $F \vdash Y \rightarrow Z$ folgt $F \vdash X \rightarrow Z$.

Beispiel

Betrachte:

$$U := \{\text{Autor, Jahrgang, Titel}\}$$
$$F_1 := \{\text{Autor} \rightarrow \text{Jahrgang}\} .$$

Somit gilt mit Regel *Elemente von F*

$$F_1 \vdash \text{Autor} \rightarrow \text{Jahrgang} .$$

Dann folgt mit Regel *Augmentation* (für $Z = \{\text{Autor, Titel}\}$)

$$F_1 \vdash \{\text{Autor, Titel}\} \rightarrow \{\text{Autor, Jahrgang, Titel}\} .$$

Theoreme

Theorem (Korrektheit und Vollständigkeit)

Ist F eine Menge von funktionalen Abhängigkeiten über U , so gilt für alle $X, Y \subseteq U$

$$F \vdash X \rightarrow Y \iff F \models X \rightarrow Y .$$

Theorem (Charakterisierung der Hülle F^+)

Ist F eine Menge von funktionalen Abhängigkeiten über U , so gilt

$$F^+ = \{X \rightarrow Y \mid F \models X \rightarrow Y\} = \{X \rightarrow Y \mid F \vdash X \rightarrow Y\} .$$

Beispiel

Im vorherigen Beispiel haben wir im Armstrong-Kalkül

$$F_1 \vdash \{\text{Autor, Titel}\} \rightarrow \{\text{Autor, Jahrgang, Titel}\}$$

hergeleitet. Mit obigem Korollar folgt nun daraus

$$\{\text{Autor, Titel}\} \rightarrow \{\text{Autor, Jahrgang, Titel}\} \in F_1^+ .$$

Folgerungen

Theorem

Ist F eine Menge von funktionalen Abhängigkeiten über U , so gilt für alle $X, Y, Z \subseteq U$:

5. Vereinigung. $F \vdash X \rightarrow Y$ und $F \vdash X \rightarrow Z \implies F \vdash X \rightarrow YZ$,

6. Zerlegung. $F \vdash X \rightarrow Y$ und $Z \subseteq Y \implies F \vdash X \rightarrow Z$.

7. Einfachheit.

$$F \vdash X \rightarrow Y_1 Y_2 \dots Y_n \iff F \vdash X \rightarrow Y_i \text{ für alle } 1 \leq i \leq n.$$

Definition (Einfache funktionale Abhängigkeit)

Funktionale Abhängigkeiten mit einelementigen rechten Seiten werden als *einfache funktionale Abhängigkeiten* bezeichnet.

Attributhülle

Definition (Attributhülle X^+)

Gegeben seien eine Menge F von funktionalen Abhängigkeiten über der universellen Menge U sowie eine Menge $X \subseteq U$. Die *Attributhülle* X^+ von X unter F ist dann definiert durch

$$X^+ := \{A \in U \mid F \vdash X \rightarrow A\} .$$

Lemma

Lemma

Für alle Mengen F von funktionalen Abhängigkeiten über der universellen Menge U sowie für alle Mengen $X, Y \subseteq U$ gilt

$$F \vdash X \rightarrow Y \iff Y \subseteq X^+ .$$

Beweis: Es sei Y die Attributmengemenge $\{A_1, A_2, \dots, A_n\}$.

Von links nach rechts. Es gelte also $F \vdash X \rightarrow Y$. Aufgrund der Zerlegungseigenschaft folgt $F \vdash X \rightarrow A_i$ für alle $1 \leq i \leq n$. Dies ergibt $Y \subseteq X^+$.

Von rechts nach links. Sei $Y \subseteq X^+$. Mit der Definition von X^+ folgt daraus $F \vdash X \rightarrow A_i$ für alle $1 \leq i \leq n$. Mit der Vereinigungseigenschaft erhalten wir also unmittelbar $F \vdash X \rightarrow Y$.

Algorithmus zur Berechnung von X^+

Gegeben seien eine universelle Menge U von Attributen, eine Menge F von funktionalen Abhängigkeiten über U sowie eine Menge $X \subseteq U$.

Eingabe: F, X

$R := X; \quad alt_R := \emptyset;$

WHILE $R \neq alt_R$ DO

$alt_R := R;$

 FOR EACH $V \rightarrow W$ IN F DO

 IF $V \subseteq R$ THEN $R := R \cup W$

Ausgabe: $X^+ := R$

Überdeckungen

Definition

Gegeben seien Mengen F und G von funktionalen Abhängigkeiten über U .

- 1 G überdeckt F , in Zeichen $F \leq G$, falls $F^+ \subseteq G^+$.
- 2 F und G sind äquivalent, falls F und G sich gegenseitig überdecken, das heisst $F \simeq G \quad :\iff \quad F \leq G \wedge G \leq F$.

Lemma

Gegeben seien Mengen F und G von funktionalen Abhängigkeiten über U .

Dann gilt $F \subseteq G^+ \iff F \leq G$.

Beweis: Rechts nach links: offensichtlich.

Links nach rechts: Sei $F \subseteq G^+$. Damit gilt

$$\begin{aligned} F^+ &= \{Y \rightarrow Z \mid F \vdash Y \rightarrow Z\} \\ &\subseteq \{Y \rightarrow Z \mid G^+ \vdash Y \rightarrow Z\} = (G^+)^+ = G^+ . \end{aligned}$$

Daraus folgt mit der vorhergehenden Definition sofort die Behauptung, dass $F \leq G$ gilt.

Test auf $F \leq G$

Eingabe: F, G .

Überprüfe für jedes $Y \rightarrow Z \in F$, ob $Y \rightarrow Z \in G^+$ durch:

Berechne Y^+ unter G und teste $Z \subseteq Y^+$.

Falls dies immer der Fall ist,

Ausgabe: $F \leq G$; anderenfalls Ausgabe: $F \not\leq G$.

Lemma

Lemma

Jede Menge F von funktionalen Abhängigkeiten über U ist zu einer Menge G äquivalent, die nur einfache funktionale Abhängigkeiten enthält, d.h. funktionale Abhängigkeiten mit einelementigen rechten Seiten.

Beweis: Wir gehen aus von einer Menge F von funktionalen Abhängigkeiten über U und definieren

$$G_F := \{X \rightarrow A \mid \exists Y (X \rightarrow Y \in F \text{ und } A \in Y)\} .$$

Aufgrund der Zerlegungseigenschaft folgt $X \rightarrow A \in F^+$ für alle A und Y mit $A \in Y$ und $X \rightarrow Y \in F$. Folglich ist $G_F \subseteq F^+$.

Ist andererseits Y die Menge $\{A_1, A_2, \dots, A_n\}$ und gilt

$$X \rightarrow A_i \in G_F \quad \text{für alle } 1 \leq i \leq n,$$

so folgt mit der Vereinigungseigenschaft auch $X \rightarrow Y \in G_F^+$. Daraus folgt $F \subseteq G_F^+$.

$G_F \subseteq F^+$ und $F \subseteq G_F^+$ ergeben $F \simeq G_F$.

G_F ist eine Menge von einfachen funktionalen Abhängigkeiten.

Minimalität

Definition

Eine Menge F von funktionalen Abhängigkeiten über U heisst *minimal*, falls folgende Bedingungen erfüllt sind:

- 1 F enthält nur einfache funktionale Abhängigkeiten.
- 2 Keine funktionale Abhängigkeit in F ist redundant, d.h.

$$(\forall X \rightarrow A \in F)(F \setminus \{X \rightarrow A\} \neq F) . \quad (\text{M2})$$

- 3 Kein Attribut auf der linken Seite einer funktionalen Abhängigkeit aus F ist redundant, d.h.

$$(\forall X \rightarrow A \in F)(\forall Y \subsetneq X)((F \setminus \{X \rightarrow A\}) \cup \{Y \rightarrow A\} \neq F) \quad (\text{M3})$$

Definition

Gegeben sei eine Menge F von funktionalen Abhängigkeiten über U . Eine *minimale Überdeckung* von F ist eine minimale Menge von funktionalen Abhängigkeiten, welche zu F äquivalent ist.

Algorithmus für minimale Überdeckungen

Eingabe: Eine Menge F von funktionalen Abhängigkeiten über U .

- 1 Spalte alle nicht-einfachen funktionalen Abhängigkeiten in F auf. Nenne die neue Menge F' .
- 2 Entferne sukzessive alle redundanten Attribute im Sinne von (M3) der vorhergehenden Definition aus F' . Nenne die neue Menge F'' .
- 3 Entferne sukzessive alle redundanten funktionalen Abhängigkeiten im Sinne von (M2) der vorhergehenden Definition aus F'' . Nenne die neue Menge F''' .

Ausgabe: $MU(F) := F'''$.

Beispiel

Attributmenge $\{A, B, C\}$ mit $F = \{AB \rightarrow C, A \rightarrow AB, B \rightarrow A\}$.

1. Schritt: Die einzige nicht-einfache funktionale Abhängigkeit in F ist $A \rightarrow AB$. Wir spalten diese auf und erhalten so die Menge

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

2. Schritt: Wir testen, ob A redundant in $AB \rightarrow C$ ist. Dazu überprüfen wir, ob $C \in \{B\}^+$ unter F' gilt.

Wegen $\{B\}^+ = \{A, B, C\}$, ist dies der Fall. Daher ist A redundant in $AB \rightarrow C$ und wird gestrichen. Weitere redundante Attribute auf linken Seiten gibt es nicht. Also erhalten wir

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

3. Schritt: Wir entfernen alle redundanten funktionalen Abhängigkeiten aus F'' . Offensichtlich gilt $A \in \{A\}^+$ unter $F'' \setminus \{A \rightarrow A\}$.

Folglich ist $A \rightarrow A$ in F'' redundant und wird entfernt. Weitere Redundanzen gibt es dann nicht mehr. Es folgt also

$$\text{MU}(F) = \{B \rightarrow C, A \rightarrow B, B \rightarrow A\} .$$

Bemerkung

Sei F eine Menge von funktionalen Abhängigkeiten. Dann kann es mehrere minimale Überdeckungen von F geben.

Betrachte die beiden Mengen von funktionalen Abhängigkeiten über der Attributmengende $\{A, B, C\}$:

$$F_1 := \{A \rightarrow B, B \rightarrow A, A \rightarrow C, C \rightarrow A\} ,$$

$$F_2 := \{A \rightarrow B, B \rightarrow C, C \rightarrow A\} .$$

Beide Mengen sind offensichtlich minimal und es gilt $F_1^+ = F_2^+$.

Repetition

- ① jedes Schema kann verlustfrei in BCNF zerlegt werden;
- ② jedes Schema kann verlustfrei und abhängigkeiterhaltend in 3NF zerlegt werden.

Algorithmus für die BCNF-Zerlegung

Gegeben seien ein Relationenschema \mathcal{S} mit einer Menge von funktionalen Abhängigkeiten F .

Eingabe: \mathcal{S}, F

$\mathcal{Z} := \{\mathcal{S}\}$

WHILE es gibt $\mathcal{S}_i \in \mathcal{Z}$ mit \mathcal{S}_i nicht in BCNF bez. $\Pi_{\mathcal{S}_i}(F)$ DO

 wähle ein solches \mathcal{S}_i

 wähle $X \rightarrow A \in F^+$ mit

$X \subseteq \mathcal{S}_i, A \in \mathcal{S}_i, A \notin X, X \rightarrow \mathcal{S}_i \notin F^+$

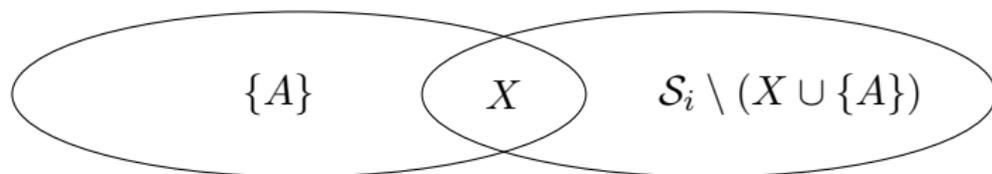
$\mathcal{Z} := (\mathcal{Z} \setminus \{\mathcal{S}_i\}) \cup \{X \cup \{A\}, \mathcal{S}_i \setminus \{A\}\}$

Ausgabe: \mathcal{Z}

Die so erhaltene Ausgabemenge \mathcal{Z} ist dann die gewünschte Zerlegung von \mathcal{S} in BCNF bezüglich F .

Bild

Wähle ein Schema \mathcal{S}_i welches die BCNF Bedingungen verletzt bezüglich einer funktionalen Abhängigkeit $X \rightarrow A$. Das Schema \mathcal{S}_i wird dann wie folgt in zwei Schemata zerlegt:



Aus $X \rightarrow A \in F^+$ folgt, dass diese Zerlegung von \mathcal{S}_i in $\{X \cup \{A\}, \mathcal{S}_i \setminus \{A\}\}$ verlustfrei ist. Dieser Schritt wird solange wiederholt, bis alle Schemata der Zerlegung in BCNF sind.

Beispiel

Wir betrachten

$$\mathcal{S}_3 = (\text{Stadt}, \text{Str}, \text{PLZ})$$

$$F_3 = \{ \{ \text{Stadt}, \text{Str} \} \rightarrow \{ \text{PLZ} \}, \{ \text{PLZ} \} \rightarrow \{ \text{Stadt} \} \} .$$

Wir beginnen nun mit $\mathcal{Z} := \{ \mathcal{S}_3 \}$. Es gilt $\mathcal{S}_3 \in \mathcal{Z}$ ist nicht in BCNF bezüglich F_3 . Wir wählen nun eine funktionale Abhängigkeit $X \rightarrow A \in F_3^+$, welche die BCNF Bedingungen verletzt. Wir setzen

$$X := \{ \text{PLZ} \} \subseteq \mathcal{S}_3 \quad \text{und} \quad A := \text{Stadt} \in \mathcal{S}_3 .$$

Es gilt somit $A \notin X$ und $X \rightarrow \mathcal{S}_3 \notin F_3^+$. Wir setzen \mathcal{Z} neu auf

$$\{ \{ \text{Stadt}, \text{PLZ} \}, \{ \text{Str}, \text{PLZ} \} \} .$$

Jetzt sind alle Elemente von \mathcal{Z} in BCNF und der Algorithmus gibt die Zerlegung \mathcal{Z} als Resultat zurück.

Algorithmus für die 3NF-Zerlegung

Gegeben Schema \mathcal{S} und *minimale* Menge F von funktionalen Abhängigkeiten. Für jede Abhängigkeit $X \rightarrow A \in F$ definieren wir ein Schema $\mathcal{S}_{X \rightarrow A}$, so dass $\mathcal{S}_{X \rightarrow A} = X \cup \{A\}$.

Eingabe: \mathcal{S}, F

$\mathcal{Z} := \{\mathcal{S}_{X \rightarrow A} \mid X \rightarrow A \in F\}$

IF kein $\mathcal{S}_{X \rightarrow A} \in \mathcal{Z}$ enthält Schlüssel für \mathcal{S} bez. F THEN

 wähle Schema \mathcal{K} , welches Schlüssel für \mathcal{S} ist

$\mathcal{Z} := \mathcal{Z} \cup \{\mathcal{K}\}$

Ausgabe: \mathcal{Z}

Die so erhaltene Ausgabemenge \mathcal{Z} ist dann die gewünschte Zerlegung von \mathcal{S} in 3NF bezüglich F .

Beispiel

Betrachte: $S_2 = (\text{BuchId}, \text{Autor}, \text{Jahrgang}, \text{Titel})$
 $F_2 = \{ \{\text{BuchId}\} \rightarrow \{\text{Autor}, \text{Jahrgang}, \text{Titel}\},$
 $\{\text{Autor}\} \rightarrow \{\text{Jahrgang}\} \} .$

Zuerst berechnen wir die minimale Überdeckung $\text{MU}(F_2)$ von F_2 .

- 1 Durch Aufspalten erhalten wir

$$F'_2 := \{ \text{BuchId} \rightarrow \text{Autor}, \text{BuchId} \rightarrow \text{Jahrgang} \\ \text{BuchId} \rightarrow \text{Titel}, \text{Autor} \rightarrow \text{Jahrgang} \} .$$

- 2 Redundante Attribute entfernen. Die linken Seiten sind alle ein-elementig. Somit haben wir $F''_2 := F'_2$.
- 3 Redundante Abhängigkeiten entfernen. Wir finden

$$\text{Jahrgang} \in \text{BuchId}^+ \text{ unter } F''_2 \setminus \{ \text{BuchId} \rightarrow \text{Jahrgang} \} .$$

Somit gilt $F''_2 \setminus \{ \text{BuchId} \rightarrow \text{Jahrgang} \} \simeq F''_2$ und wir setzen

$$\text{MU}(F_2) := F'''_2 := F''_2 \setminus \{ \text{BuchId} \rightarrow \text{Jahrgang} \} .$$

Beispiel

Als Input für den 3NF-Algorithmus verwenden wir nun \mathcal{S}_2 und $MU(F_2)$.

$$\mathcal{S}_2 = (\text{BuchId}, \text{Autor}, \text{Jahrgang}, \text{Titel})$$

$$MU(F_2) := \{ \text{BuchId} \rightarrow \text{Autor}, \text{BuchId} \rightarrow \text{Titel}, \text{Autor} \rightarrow \text{Jahrgang} \}$$

Im ersten Schritt erhalten wir so die Zerlegung

$$\mathcal{Z} = \{ \{ \text{BuchId}, \text{Autor} \}, \{ \text{BuchId}, \text{Titel} \}, \{ \text{Autor}, \text{Jahrgang} \} \} .$$

Das Schema $\{ \text{BuchId}, \text{Autor} \} \in \mathcal{Z}$ enthält einen Schlüssel für \mathcal{S}_2 bezüglich $MU(F_2)$. Somit müssen wir im zweiten Schritt kein Schema hinzufügen.

Damit ist \mathcal{Z} eine verlustfreie und abhängigkeiterhaltende Zerlegung des Schemas \mathcal{S}_2 in die dritte Normalform.