Relationale Entwurfstheorie

Funktionale Abhängigkeiten
Normalformen
Normalisierung durch Dekomposition

Vorlesung “Informationssysteme” SS 2007
Dan Olteanu, Lehrstuhl für Informationssysteme, Universität des Saarlandes

Relationale Entwurfstheorie

Das relationale Modell kann die Bedeutung der Daten nicht bestimmen.
Die relationale Entwurfstheorie bietet eine Möglichkeit an, die Anwendungssemantik der Daten durch Konsistenzbedingungen zu spezifizieren.

Ziele der relationalen Entwurfstheorie:
Bewertung der Qualität eines Relationenschemas

- Redundanz und Anomalien (also "schlechte" Relationenschemata) vermeiden
- Einhaltung von Konsistenzbedingungen (z.B. funktionaler Abhängigkeiten)
  Man kann existierende Konsistenzbedingungen auch für die Anfrage-Optimierung benutzen (wird später in der Vorlesung betrachtet).

Normalformen als Gütekriterium

Verbesserung eines Relationenschemas

- Durch den Synthesealgorithmus
- Durch Dekomposition

“Schlechte” Relationenschemata

ProfVorl

PersNr	Name	Rang	Raum	VorlNr	Titel	SWS
2125	Sokrates	C4	226	5041	Ethik	4
2125	Sokrates	C4	226	5049	Mäeutik	2
2125	Sokrates	C4	226	4052	Logik	4
...	...	...	...	...	...	...
2132	Popper	C3	52	5259	Der Wiener Kreis	2
2137	Kant	C4	7	4630	Die 3 Kritiken	4

- Update-Anomalien
  Sokrates zieht um, von Raum 226 in Raum 338. Was passiert?
- Einfüge-Anomalien
  Neue/r Prof ohne Vorlesungen?
- Löschanomalien
  Letzte Vorlesung einer/s Prof wird gelöscht. Was passiert?

Ein besserer Entwurf: Zerlege ProfVorl in 2 Relationen – Professoren und Vorlesungen.

Schlüssel und Abhängigkeiten

PersNr soll ein Schlüssel für die Relation ProfVorl sein, d.h.
es gibt keine zwei Professoren mit demselben Schlüssel PersNr.

Die Existenz eines Schlüssels ist relevant für

- Speicherung und Anfrageauswertung (Indizes, Umformulierung)
- die Anwendungssemantik und Konsistenz der Daten

Funktionale Abhängigkeiten (FDs)

Seien

- ein Schema $\mathcal{R}$ und zwei Teilmengen $\alpha, \beta \subseteq \mathcal{R}$
- eine Ausprägung R von $\mathcal{R}$.

$$\boxed{\text{FD } \alpha \rightarrow \beta \text{ genau dann wenn } \forall r, s \in R : r.\alpha = s.\alpha \Rightarrow r.\beta = s.\beta}$$

Bemerkung: $r.\alpha = s.\alpha$ ist eine Kurzform für $\forall A \in \alpha : r.A = s.A$.
$\alpha \rightarrow \beta$ bedeutet dass

- die α-Werte die β-Werte funktional bestimmen, oder
- die β-Werte funktional abhängig von den α-Werten sind, oder (kürzer)
- β funktional abhängig von α ist.

FD Beispiel 1

R	A	B	C	D
	a_4	b_2	c_4	d_3
	a_1	b_1	c_1	d_1
	a_1	b_1	c_1	d_2
	a_2	b_2	c_3	d_2
	a_3	b_2	c_4	d_3

Gilt FD $\{A\} \rightarrow \{B\}$, d.h. $\forall r, s \in R : r.A = s.A \Rightarrow r.B = s.B$?

FD Beispiel 1

R	A	B	C	D
	a_4	b_2	c_4	d_3
$*$	a_1	b_1	c_1	d_1
$*$	a_1	b_1	c_1	d_2
	a_2	b_2	c_3	d_2
	a_3	b_2	c_4	d_3

Gilt FD $\{A\} \rightarrow \{B\}$, d.h. $\forall r, s \in R : r.A = s.A \Rightarrow r.B = s.B$?
Ja. Begründung:

- Die mit $*$ markierten Tupel haben dengleichen A-Wert a_1.
- Es stimmt auch, dass sie dengleichen B-Wert b_1 haben.

FD Beispiel 2

R	A	B	C	D
	a_4	b_2	c_4	d_3
	a_1	b_1	c_1	d_1
	a_1	b_1	c_1	d_2
	a_2	b_2	c_3	d_2
	a_3	b_2	c_4	d_3

Gilt FD $\{B\} \rightarrow \{C\}$, d.h. $\forall r, s \in R : r.B = s.B \Rightarrow r.C = s.C$?

FD Beispiel 2

R	A	B	C	D
$*$	a_4	$\mathbf{b_2}$	c_4	d_3
$**$	a_1	b_1	c_1	d_1
$**$	a_1	b_1	c_1	d_2
$*$	a_2	$\mathbf{b_2}$	c_3	d_2
$*$	a_3	b_2	c_4	d_3

Gilt FD $\{B\} \rightarrow \{C\}$, d.h. $\forall r, s \in R : r.B = s.B \Rightarrow r.C = s.C$?
Nein. Begründung:

- Die mit $*$ markierten Tupel haben dengleichen B-Wert b_2
- aber unterschiedlichen C-Werte (c_3 und c_4).

FD Beispiel 3

R	A	B	C	D
	a_4	b_2	c_4	d_3
	a_1	b_1	c_1	d_1
	a_1	b_1	c_1	d_2
	a_2	b_2	c_3	d_2
	a_3	b_2	c_4	d_3

Gilt FD $\{A, D\} \rightarrow \{B\}$, d.h., $\forall r, s \in R : r.A = s.A \wedge r.D = s.D \Rightarrow r.B = s.B$?

FD Beispiel 3

R	A	B	C	D
	a_4	b_2	c_4	d_3
	a_1	b_1	c_1	d_1
	a_1	b_1	c_1	d_2
	a_2	b_2	c_3	d_2
	a_3	b_2	c_4	d_3

Gilt FD $\{A, D\} \rightarrow \{B\}$, d.h., $\forall r, s \in R : r.A = s.A \wedge r.D = s.D \Rightarrow r.B = s.B$?
Ja. Begründung:

- In R existieren keine unterschiedlichen Tupel mit dengleichen A- und D-Werten.

Weiteres Beispiel: FD $\{C, D\} \rightarrow \{B\}$ gilt.

Volle funktionale Abhängigkeit

Schema $\mathcal{R}$ und $\alpha, \beta \subseteq \mathcal{R}$.
β ist *voll funktional abhängig* von α (geschrieben $\alpha \stackrel{\bullet}{\rightarrow} \beta$) genau dann wenn

- $\alpha \rightarrow \beta$
- α kann nicht mehr "verkleinert" werden $\qquad \forall A \in \alpha : \alpha - \{A\} \not\rightarrow \beta$

Unser Beispiel-Schema mit FDs $\{A, D\} \rightarrow \{B\}$ und $\{A\} \rightarrow \{B\}$.

R	A	B	C	D
	a_4	b_2	c_4	d_3
	a_1	b_1	c_1	d_1
	a_1	b_1	c_1	d_2
	a_2	b_2	c_3	d_2
	a_3	b_2	c_4	d_3

Da $\{A\} \subset \{A, D\}$, FD $\{A, D\} \stackrel{\bullet}{\rightarrow} \{B\}$ gilt nicht, aber FD $\{A\} \stackrel{\bullet}{\rightarrow} \{B\}$ gilt.

Schlüssel

Schema $\mathcal{R}$ und $\alpha \subseteq \mathcal{R}$.

$$\boxed{\alpha \text{ ist ein } \textit{Superschlüssel} \text{ falls } \alpha \to \mathcal{R} \text{ gilt.}}$$

D.h. α bestimmt alle anderen Attributwerte einer $\mathcal{R}$-Ausprägung.
Bemerkung: $\mathcal{R} \to \mathcal{R}$, d.h. die Menge aller Attribute eines Schema bildet einen Superschlüssel.

$$\boxed{\alpha \text{ ist ein } \textit{Kandidatenschlüssel} \text{ falls } \alpha \xrightarrow{\bullet} \mathcal{R} \text{ gilt.}}$$

Ein *Primärschlüssel* ist ein Kandidatenschlüssel.

- Auswahl des Primärschlüssels wichtig für die Verweise zwischen Tupeln unterschiedlicher Relationen.
- Der Primärschlüssel kann *Fremdschlüssel* für andere Relationen sein.

Beispiel:

- Relation R über $\{A, B\}$ mit Primärschlüssel $\{A\}$,
- Relationen S über $\{C, A\}$ und T über $\{D, A\}$, beide mit Fremdschlüssel $\{A\}$.

Herleitung funktionaler Abhängigkeiten

Armstrong-Axiome (Herleitungsregeln)

1. **Reflexivität**: $\beta \subseteq \alpha \Rightarrow \alpha \to \beta$
2. **Verstärkung**: $\alpha \to \beta \Rightarrow \alpha\gamma \to \beta\gamma$ ($\alpha\gamma$ bedeutet $\alpha \cup \gamma$)
3. **Transitivität**: $\alpha \to \beta, \beta \to \gamma \Rightarrow \alpha \to \gamma$

Diese drei Axiome sind korrekt und vollständig, d.h.

- die aus einer FD-Menge F herleitbaren FDs sind von *jeder* Ausprägung erfüllt, für die F auch erfüllt ist.
- *alle* FDs F^+ sind aus einer FD-Menge F herleitbar, die durch F logisch impliziert werden.

$$\boxed{\text{Zu einer Menge von FDs } F \text{ gibt es eine eindeutige Hülle } F^+}$$

Beispiel: FDs $\{\{A, B\} \to \{B, C\}, CD \to \{E, F\}, \{B, E\} \to \{A, D\}, \{C\} \to \{D\}\}$.

- $\{C\} \to \{C\}$ (Reflexivität)
- $\{C\} \to \{D\} \Rightarrow \{A, C\} \to \{A, D\}$ (Verstärkung)
- $\{C\} \to \{D\} \Rightarrow \{C\} \to \{C, D\} \Rightarrow \{C\} \to \{E, F\}$ (Verstärkung,Transitivität)

Zusätzliche Axiome zur Erleichterung der Herleitung

1. **Pseudotransitivitätsregel**: $\alpha \to \beta, \gamma\beta \to \delta \Rightarrow \alpha\gamma \to \delta$
   $\alpha \to \beta \overset{\text{Verst.}}{\Longrightarrow} \alpha\gamma \to \beta\gamma \overset{\text{Trans.}}{\Longrightarrow} \alpha\gamma \to \delta$.

2. **Dekompositionsregel**: $\alpha \to \beta\gamma \Rightarrow \alpha \to \beta, \alpha \to \gamma$
   Durch Reflexivität, $\beta\gamma \to \beta$ und $\beta\gamma \to \gamma$. Dann, $\alpha \to \beta\gamma \overset{\text{Trans.}}{\Longrightarrow} \alpha \to \beta$.

3. **Vereinigungsregel**: $\alpha \to \beta, \alpha \to \gamma \Rightarrow \alpha \to \beta\gamma$
   $\alpha \to \beta \overset{\text{Verst.}}{\Longrightarrow} \alpha\gamma \to \beta\gamma$. Auch $\alpha \to \gamma \overset{\text{Verst.}}{\Longrightarrow} \alpha \to \alpha\gamma$.
   Dann durch Transitivität, $\alpha \to \alpha\gamma \to \beta\gamma$.

Bestimmung der Hülle einer Attributmenge

Eingabe: Schema $\mathcal{R}$, eine Menge F von FDs und eine Menge von Attributen $\alpha \subseteq \mathcal{R}$.
Ausgabe: Die vollständige Menge von Attributen α^+, für die gilt $\alpha \rightarrow \alpha^+$.

<u>AttrHülle</u> (F, α)

  Ergebnis $:= \alpha$
  While (Änderungen an Ergebnis) **do**
    Foreach FD $\beta \rightarrow \gamma$ in F **do**
      If $\beta \subseteq$ Ergebnis **then** Ergebnis $:=$ Ergebnis $\cup \gamma$
  Return $\alpha^+ =$ Ergebnis

$\boxed{\alpha \text{ ist ein Superschlüssel genau dann wenn AttrHülle}(F, \alpha) = \mathcal{R}.}$

AttrHülle und Schlüssel: Beispiel 1

Schema $\mathcal{R} = \{A, B, C, D, E, F\}$.
FDs $\{\{A, B\} \to \{B, C\}, \{C, D\} \to \{E, F\}, \{B, E\} \to \{A, D\}, \{C\} \to \{D\}\}$.

AttrHülle (FDs, X) $= X$, für $X \in \{\{A\}, \{B\}, \{D\}, \{E\}, \{F\}\}$.

AttrHülle (FDs, $\{C\}$) $= \{C, D, E, F\}$.
AttrHülle (FDs, $\{A, B\}$) $= \{A, B, C, D, E, F\}$. $\{A, B\}$ ist ein Superschlüssel.
AttrHülle (FDs, $\{B, C\}$) $= \{B, C, D, E, F, A, \}$. $\{B, C\}$ ist ein Superschlüssel.

$\{A, B\}$ und $\{B, C\}$ sind auch Kandidatenschlüssel, weil $\{A\}$, $\{B\}$ oder $\{C\}$ keine Kandidatenschlüssel sind.

AttrHülle und Schlüssel: Beispiel 2

Schema $\mathcal{R} = \{A, B, C, D\}$. FDs $\{\{A, B\} \to \{C\}, \{C\} \to \{D\}, \{D\} \to \{A\}\}$.

Superschlüssel: $\{A, B\}, \{B, C\}, \{B, D\}$.

AttrHülle (FDs, $\{A, B\}$)
$= \{A, B\}$ (Reflexivität)
$= \{A, B, C\}$ (weil $\{A, B\} \to \{C\}$)
$= \{A, B, C, D\}$ (weil $\{C\} \to \{D\}$).

AttrHülle (FDs, $\{B, C\}$)
$= \{B, C\}$ (Reflexivität)
$= \{B, C, D\}$ (weil $\{C\} \to \{D\}$)
$= \{B, C, D, A\}$ (weil $\{D\} \to \{A\}$).

AttrHülle (FDs, $\{B, D\}$)
$= \{B, D\}$ (Reflexivität)
$= \{B, D, A\}$ (weil $\{D\} \to \{A\}$)
$= \{B, D, A, C\}$ (weil $\{A, B\} \to \{C\}$).

Kandidatenschlüssel: $\{A, B\}, \{B, C\}, \{B, D\}$.

Äquivalenz der Mengen von FDs

Schema $\mathcal{R}$, zwei Mengen von FDs F und G und deren Hüllen F^+ und G^+.

$$\boxed{F \text{ und } G \text{ sind äquivalent, geschrieben } F \equiv G \text{, genau dann wenn } F^+ = G^+}$$

Bemerkung 1: Zu einer Menge von FDs F gibt es eine eindeutige Hülle F^+.
Bemerkung 2: F^+ könnte sehr viele (teilweise redundanten) FDs enthalten.

Beispiel:

- $\mathcal{R} = \{A, B, C, D, E, F\}$,
- $F = \{\{A, B\} \to \{B, C\}, \{C, D\} \to \{E, F\}, \{B, E\} \to \{A, D\}, \{C\} \to \{D\}\}$,
- $G = \{\{A, B\} \to \{C\}, \{C\} \to \{D, E, F\}, \{B, E\} \to \{A\}\}$
- $G \neq F$ und $G \equiv F$.
  Hier G ist eine kleinstmögliche noch äquivalente Menge von FDs.

Kanonische Überdeckung von FD-Mengen

Menge F von FDs. Die kanonische Überdeckung von F ist *eine* kleinstmögliche noch äquivalente Menge Fc von FDs:

1. $Fc \equiv F$, d.h. $Fc^+ = F^+$
2. In Fc existieren keine FDs, die überflüssige Attribute enthalten. D.h. es muss folgendes gelten:
   - [Linksreduktion] $\forall A \in \alpha : (Fc - \{(\alpha \rightarrow \beta)\} \cup \{((\alpha - \{A\}) \rightarrow \beta)\}) \not\equiv Fc$
   - [Rechtsreduktion] $\forall B \in \beta : (Fc - \{(\alpha \rightarrow \beta)\} \cup \{(\alpha \rightarrow (\beta - \{B\}))\}) \not\equiv Fc$
3. Jede linke Seite einer funktionalen Abhängigkeit in Fc ist einzigartig. Dies kann durch sukzessive Anwendung der Vereinigungsregel auf FDs der Art $\alpha \rightarrow \beta$ und $\alpha \rightarrow \gamma$ erzielt werden, so dass die beiden FDs durch $\alpha \rightarrow \beta\gamma$ ersetzt werden.

Berechnung der kanonischen Überdeckung

1. [Linksreduktion] Für jede FD $\alpha \to \beta$, für alle $A \in \alpha$ überprüfe ob A redundant ist, d.h. $\beta \subseteq \text{AttrHülle}(F, \alpha - \{A\})$. Falls dies der Fall ist, ersetze $\alpha \to \beta$ durch $\alpha - \{A\} \to \beta$.

2. [Rechtsreduktion] Für jede FD $\alpha \to \beta$, für alle $B \in \beta$ überprüfe ob B redundant ist, d.h. $B \in \text{AttrHülle}(F - (\alpha \to \beta) \cup \{(\alpha \to (\beta - \{B\}))\}, \alpha)$. Falls dies der Fall ist, ersetze $\alpha \to \beta$ durch $\alpha \to \beta - \{B\}$.

3. Entferne die FDs der Form $\alpha \to \emptyset$, die im 2. Schritt möglicherweise entstanden sind.

4. Fasse mittels der Vereinigungsregel FDs der Form $\alpha \to \beta_1, \ldots, \alpha \to \beta_n$ zusammen, so dass $\alpha \to \beta_1 \ldots \beta_n$ verbleibt.

Kanonische Überdeckung: Beispiel

Schema $\mathcal{R} = \{A, B, C, D, E, F\}$.
FDs $\{\{A, B\} \to \{B, C\}, \{C, D\} \to \{E, F\}, \{B, E\} \to \{A, D\}, \{C\} \to \{D\}\}$.

1. Linksreduktion.
AttrHülle (FDs, X) $= X$, für $X \in \{\{A\}, \{B\}, \{D\}, \{E\}, \{F\}\}$.
AttrHülle (FDs, $\{C\}$) $= \{C, D, E, F\} \supset \{E, F\}$.
Ersetze also $\{C, D\} \to \{E, F\}$ durch $\{C\} \to \{E, F\}$.

Kanonische Überdeckung: Beispiel

Schema $\mathcal{R} = \{A, B, C, D, E, F\}$.
FDs $\{\{A, B\} \to \{B, C\}, \{C\} \to \{E, F\}, \{B, E\} \to \{A, D\}, \{C\} \to \{D\}\}$.

1. Linksreduktion.
AttrHülle (FDs, X) $= X$, für $X \in \{\{A\}, \{B\}, \{D\}, \{E\}, \{F\}\}$.
AttrHülle (FDs, $\{C\}$) $= \{C, D, E, F\} \supset \{E, F\}$.
Ersetze also $\{C, D\} \to \{E, F\}$ durch $\{C\} \to \{E, F\}$.

Kanonische Überdeckung: Beispiel

Schema $\mathcal{R} = \{A, B, C, D, E, F\}$.
FDs $\{\{A, B\} \to \{B, C\}, \{C\} \to \{E, F\}, \{B, E\} \to \{A, D\}, \{C\} \to \{D\}\}$.

2. Rechtsreduktion.
$B \in$ AttrHülle (FDs, $\{A, B\}$).
Ersetze also $\{A, B\} \to \{B, C\}$ durch $\{A, B\} \to \{C\}$.

Kanonische Überdeckung: Beispiel

Schema $\mathcal{R} = \{A, B, C, D, E, F\}$.
FDs $\{\{A, B\} \rightarrow \{C\}, \{C\} \rightarrow \{E, F\}, \{B, E\} \rightarrow \{A, D\}, \{C\} \rightarrow \{D\}\}$.

2. Rechtsreduktion.
$B \in$ AttrHülle (FDs, $\{A, B\}$).
Ersetze also $\{A, B\} \rightarrow \{B, C\}$ durch $\{A, B\} \rightarrow \{C\}$.

Kanonische Überdeckung: Beispiel

Schema $\mathcal{R} = \{A, B, C, D, E, F\}$.
FDs $\{\{A, B\} \rightarrow \{C\}, \{C\} \rightarrow \{E, F\}, \{B, E\} \rightarrow \{A, D\}, \{C\} \rightarrow \{D\}\}$.

2. Rechtsreduktion.
$D \in$ AttrHülle (FDs, $\{B, E\}$).
Ersetze also $\{B, E\} \rightarrow \{A, D\}$ durch $\{B, E\} \rightarrow \{A\}$.

Kanonische Überdeckung: Beispiel

Schema $\mathcal{R} = \{A, B, C, D, E, F\}$.
FDs $\{\{A, B\} \rightarrow \{C\}, \{C\} \rightarrow \{E, F\}, \{B, E\} \rightarrow \{A\}, \{C\} \rightarrow \{D\}\}$.

2. Rechtsreduktion.
$D \in$ AttrHülle (FDs, $\{B, E\}$).
Ersetze also $\{B, E\} \rightarrow \{A, D\}$ durch $\{B, E\} \rightarrow \{A\}$.

Kanonische Überdeckung: Beispiel

Schema $\mathcal{R} = \{A, B, C, D, E, F\}$.
FDs $\{\{A, B\} \to \{C\}, \{C\} \to \{E, F\}, \{B, E\} \to \{A\}, \{C\} \to \{D\}\}$.

3. Entferne FDs der Form $\alpha \to \emptyset$. Nichts zu tun.
4. Wende Vereinigungsregel an.

Ersetze $\{C\} \to \{E, F\}$ und $\{C\} \to \{D\}$ durch $\{C\} \to \{D, E, F\}$.

Kanonische Überdeckung: Beispiel

Schema $\mathcal{R} = \{A, B, C, D, E, F\}$.
Ursprüngliche Menge von FDs:
FDs $\{\{A, B\} \to \{B, C\}, \{C, D\} \to \{E, F\}, \{B, E\} \to \{A, D\}, \{C\} \to \{D\}\}$.

Die kanonische Überdeckung:
FDs $\{\{A, B\} \to \{C\}, \{C\} \to \{D, E, F\}, \{B, E\} \to \{A\}\}$.

Eindeutigkeit der kanonischen Überdeckung

Eine Menge von FDs kann mehrere kanonischen Überdeckungen haben!

Beispiel:
Für FDs $\{\{A\} \rightarrow \{B, C\}, \{B\} \rightarrow \{A, C\}\}$ gibt's zwei mögliche kanonische Überdeckungen:

1. $\{\{A\} \rightarrow \{B, C\}, \{B\} \rightarrow \{A\}\}$
2. $\{\{B\} \rightarrow \{A, C\}, \{A\} \rightarrow \{B\}\}$

Zerlegung (Dekomposition) von Schemata

"Schlechte Schemata" - nicht "zusammenpassende" Informationen werden in einem Schema verbündelt.
Problem: (Update-,Einfüge- und Lösch-)Anomalien und Redundanz
Lösung: Zerlegungen von Schemata (und deren Ausprägungen).

Schemata $\mathcal{R}$, $\mathcal{R}_1, \ldots, \mathcal{R}_n$.

$\boxed{\mathcal{R}_1, \ldots, \mathcal{R}_n \text{ stellen eine Zerlegung von } \mathcal{R} \text{ dar, falls } \forall 1 \leq i \leq n : \mathcal{R}_i \subseteq \mathcal{R}}$

Es gibt zwei Korrektheitskriterien für die Zerlegung von Schemata:

1. Verlustlosigkeit
   Die in der ursprünglichen Ausprägung R des Schemas $\mathcal{R}$ enthaltenen Informationen müssen aus den Ausprägungen $R_1, \ldots, R_n$ der neuen Schemata $\mathcal{R}_1, \ldots, \mathcal{R}_n$ rekonstruierbar sein.
2. Abhängigkeitserhaltung
   Die für $\mathcal{R}$ geltenden funktionalen Abhängigkeiten müssen auf die Schemata $\mathcal{R}_1, \ldots, \mathcal{R}_n$ übertragbar sein.

Verlustlosigkeit einer Zerlegung

Es reicht aus, sich auf die Zerlegung von Schema $\mathcal{R}$ in 2 Schemata $\mathcal{R}_1$ und $\mathcal{R}_2$ zu beschränken: $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$, $R_1 = \Pi_{\mathcal{R}_1}(R)$, $R_2 = \Pi_{\mathcal{R}_2}(R)$.

1. Für jede mögliche (gültige) Ausprägung R von $\mathcal{R}$ gilt $R = R_1 \bowtie R_2$.
2. Hinreichende Bedingung für die Verlustlosigkeit einer Zerlegung:

$$\boxed{\mathcal{R}_1 \cap \mathcal{R}_2 \rightarrow \mathcal{R}_1 \text{ oder } \mathcal{R}_1 \cap \mathcal{R}_2 \rightarrow \mathcal{R}_2}$$

"Verlustige" Zerlegung

Biertrinker

Kneipe	Gast	Bier
Kowalski	Kemper	Pils
Kowalski	Eickler	Hefeweizen
Innsteg	Kemper	Hefeweizen

$\Pi_{\text{Kneipe,Gast}}(\text{Biertrinker})$

Kneipe	Gast
Kowalski	Kemper
Kowalski	Eickler
Innsteg	Kemper

$\Pi_{\text{Gast,Bier}}(\text{Biertrinker})$

Gast	Bier
Kemper	Pilz
Eickler	Hefeweizen
Kemper	Hefeweizen

$\Pi_{\text{Kneipe,Gast}}(\text{Biertrinker}) \bowtie \Pi_{\text{Gast,Bier}}(\text{Biertrinker})$

Kneipe	Gast	Bier
Kowalski	Kemper	Pils
Kowalski	Kemper	Hefeweizen
Kowalski	Eickler	Hefeweizen
Innsteg	Kemper	Pils
Innsteg	Kemper	Hefeweizen

$\boxed{\text{Biertrinker} \neq \Pi_{\text{Kneipe,Gast}}(\text{Biertrinker}) \bowtie \Pi_{\text{Gast,Bier}}(\text{Biertrinker})}$

Verlustfreie Zerlegung

Eltern

Vater	Mutter	Kind
Johann	Martha	Else
Johann	Maria	Theo
Heinz	Martha	Cleo

$\Pi_{\text{Vater,Kind}}(\text{Eltern})$

Vater	Kind
Johann	Else
Johann	Theo
Heinz	Cleo

$\Pi_{\text{Mutter,Kind}}(\text{Eltern})$

Mutter	Kind
Martha	Else
Maria	Theo
Martha	Cleo

$\Pi_{\text{Vater,Kind}}(\text{Eltern}) \bowtie \Pi_{\text{Mutter,Kind}}(\text{Eltern})$

Vater	Mutter	Kind
Johann	Martha	Else
Johann	Maria	Theo
Heinz	Martha	Cleo

$$\boxed{\text{Eltern} = \Pi_{\text{Vater,Kind}}(\text{Eltern}) \bowtie \Pi_{\text{Mutter,Kind}}(\text{Eltern})}$$

(Zusätzliche) Erklärung der Beispiele

1. Biertrinker hat FD $\{\text{Kneipe,Gast}\} \rightarrow \{\text{Bier}\}$
   Bedingung $\{\text{Kneipe,Gast}\} \cap \{\text{Gast,Bier}\} = \{\text{Gast}\} \rightarrow \{\text{Kneipe,Gast}\}$ oder $\{\text{Gast}\} \rightarrow \{\text{Gast,Bier}\}$ nicht erfüllt.

2. Eltern hat FDs $\{\text{Kind}\} \rightarrow \{\text{Mutter}\}$, $\{\text{Kind}\} \rightarrow \{\text{Vater}\}$
   Bedingung $\{\text{Vater,Kind}\} \cap \{\text{Mutter,Kind}\} = \{\text{Kind}\} \rightarrow \{\text{Vater,Kind}\}$ oder $\{\text{Kind}\} \rightarrow \{\text{Mutter,Kind}\}$ erfüllt.
   $\{\text{Kind}\}$ ist auch der Schlüssel.

Kriterium der Verlustlosigkeit

Zerlegung von $\mathcal{R}$ mit FDs F in 2 Schemata $\mathcal{R}_1$ und $\mathcal{R}_2$.
Es sollte gelten (hinreichende aber nicht notwendige Bedingung):

$$\boxed{\mathcal{R}_1 \cap \mathcal{R}_2 \rightarrow \mathcal{R}_1 \text{ oder } \mathcal{R}_1 \cap \mathcal{R}_2 \rightarrow \mathcal{R}_2}$$

Das ist äquivalent zu

$$(\mathcal{R}_1 \cap \mathcal{R}_2 \rightarrow \mathcal{R}_1) \in F^+ \text{ oder } (\mathcal{R}_1 \cap \mathcal{R}_2 \rightarrow \mathcal{R}_2)) \in F^+$$

Und auch äquivalent zu
$\mathcal{R} = \alpha \cup \beta \cup \gamma, \mathcal{R}_1 = \alpha \cup \beta, \mathcal{R}_2 = \alpha \cup \gamma$

$$\boxed{\beta \subseteq \text{AttrHülle}(F, \alpha) \text{ oder } \gamma \subseteq \text{AttrHülle}(F, \alpha)}$$

Abhängigkeitsbewahrung einer Zerlegung

Zerlegung von $\mathcal{R}$ mit FDs $F_{\mathcal{R}}$ in Schemata $\mathcal{R}_1, \ldots, \mathcal{R}_n$ mit FDs $F_{\mathcal{R}_1}, \ldots, F_{\mathcal{R}_n}$.

Die Überprüfung aller FDs soll **lokal** auf $\mathcal{R}$ erfolgen, ohne die Ausprägung von $\mathcal{R}$ zu rekonstruieren (zu aufwändig).

Bedingung: Hüllentreue Dekomposition

$$\boxed{F_{\mathcal{R}} \equiv F_{\mathcal{R}_1} \cup \ldots \cup F_{\mathcal{R}_n} \text{ bzw. } F_{\mathcal{R}}^{+} = (F_{\mathcal{R}_1} \cup \ldots \cup F_{\mathcal{R}_n})^{+}}$$

Verlustlose und nicht abhängigkeitserhaltende Zerlegung

Schema PLZverzeichnis: {[Strasse, Ort, Bland, PLZ]}

Annahmen

- Orte werden durch ihren Namen (Ort) und das Bundesland (Bland) eindeutig identifiziert
- Innerhalb einer Strasse ändert sich die Postleitzahl nicht
- Postleitzahlengebiete gehen nicht über Ortsgrenzen und Orte nicht über Bundeslandgrenzen hinweg

Daraus resultieren die FDs

- $\{\text{PLZ}\} \rightarrow \{\text{ Ort, BLand}\}$
- $\{\text{Strasse, Ort, BLand}\} \rightarrow \{\text{PLZ}\}$

Betrachte die Zerlegung

- Strassen: {[PLZ, Strasse]}
- Orte: {[PLZ, Ort, BLand]}

Zerlegung des Schemas PLZVerzeichnis

PLZVerzeichnis

Ort	**Bland**	**Strasse**	PLZ
Frankfurt	Hessen	Goethestrasse	60313
Frankfurt	Hessen	Galgenstrasse	60437
Frankfurt	Brandenburg	Goethestrasse	15234

$\Pi_{\text{PLZ,Strasse}}(\text{PLZVerzeichnis})$

PLZ	**Strasse**
15234	Goethestrasse
60313	Goethestrasse
60437	Galgenstrasse

$\Pi_{\text{Ort,BLand,PLZ}}(\text{PLZVerzeichnis})$

Ort	BLand	**PLZ**
Frankfurt	Hessen	60313
Frankfurt	Hessen	60437
Frankfurt	Brandenburg	15234

Die FD $\{\text{Strasse, Ort, BLand}\} \rightarrow \{\text{PLZ}\}$ ist im zerlegten Schema nicht mehr enthalten und das Einfügen inkonsistenter Tupel ist möglich.

Einfügen zweier Tupel, die eine globale FD verletzen

PLZVerzeichnis

Ort	**Bland**	**Strasse**	PLZ
Frankfurt	Hessen	Goethestrasse	60313
Frankfurt	Hessen	Galgenstrasse	60437
Frankfurt	Brandenburg	Goethestrasse	15234

$\Pi_{PLZ,Strasse}(PLZVerzeichnis)$

PLZ	**Strasse**
15234	Goethestrasse
60313	Goethestrasse
60437	Galgenstrasse
15235	Goethestrasse

$\Pi_{Ort,BLand,PLZ}(PLZVerzeichnis)$

Ort	BLand	**PLZ**
Frankfurt	Hessen	60313
Frankfurt	Hessen	60437
Frankfurt	Brandenburg	15234
Frankfurt	Brandenburg	15235

Das Tupel (Frankfurt, Brandenburg, Goethestrasse, 15235) verletzt die globale FD $\{$Strasse, Ort, BLand$\} \rightarrow \{$PLZ$\}$.

Erste Normalform (1NF)

- von allen bis jetzt vorgestellten Relationen automatisch eingehalten.
- alle Attribute haben atomare Wertebereiche
  also zusammengesetzte, mengenwertige, oder relationenwertige Domänen nicht zulässig

Eltern

Vater	Mutter	Kinder
Johann	Martha	{Else,Lucy}
Johann	Maria	{Theo,Josef}
Heinz	Martha	{Cleo}

- Die obige Relation in 1NF

Eltern

Vater	Mutter	Kind
Johann	Martha	Else
Johann	Martha	Lucy
Johann	Maria	Theo
Johann	Maria	Josef
Heinz	Martha	Cleo

Exkurs: NF^2-Relationen

- Non-First Normal-Form Relationen
- Geschachtelte Relationen

Eltern

Vater	Mutter	Kinder	
		Name	Alter
Johann	Martha	Else	5
		Lucy	3
Johann	Maria	Theo	3
		Josef	1
Heinz	Martha	Cleo	9

Jedes Tupel der Relation Eltern enthält eine Relation Kinder.

Dritte Normalform (3NF)

Schema $\mathcal{R}$ und FDs F.
$\mathcal{R}$ ist in 3NF, wenn für jede FD $(\alpha \to B) \in F^+$ mit $\alpha \subseteq \mathcal{R}$ und $B \in \mathcal{R}$ *eine* von 3 Bedingungen gilt:

- $B \in \alpha$, d.h. die FD ist trivial
- B ist in einem Kandidatenschlüssel von $\mathcal{R}$ enthalten
- α ist Superschlüssel von $\mathcal{R}$

Die Verletzung der Normalform könnte zu Redundanz führen.

Beispiel:
Vorlesung: {[Sitzung,Prof,Zeit]}
FDs { {Sitzung} $\to$ {Prof}, {Prof,Zeit} $\to$ {Sitzung} }
{Prof,Zeit} ist ein Kandidatenschlüssel, der {Prof} enthält. Also Vorlesung ist in 3NF.

Synthesealgorithmus - Eigenschaften der Zerlegung

Schema $\mathcal{R}$ und FDs F. Zerlegung in $\mathcal{R}_1, \ldots, \mathcal{R}_n$ so dass:

- Die Zerlegung $\mathcal{R}_1, \ldots, \mathcal{R}_n$ ist **verlustlos**
- Die Zerlegung $\mathcal{R}_1, \ldots, \mathcal{R}_n$ ist **abhängigkeitsbewahrend**
- alle $\mathcal{R}_1, \ldots, \mathcal{R}_n$ sind in **3NF**

Synthesealgorithmus

1. Bestimme die kanonische Überdeckung Fc zu F. Wiederholung:
   - Links- und Rechtsreduktion
   - Entfernung von FDs der Form $\alpha \to \emptyset$
   - Zusammenfassung gleicher linker Seiten
2. Für jede FD $(\alpha \to \beta) \in Fc$:
   - Kreiere ein Schema $\mathcal{R}\alpha := \alpha \cup \beta$
   - Ordne $\mathcal{R}\alpha$ die FDs $F\alpha := \{(\alpha' \to \beta') \in Fc \mid \alpha' \cup \beta' \subseteq \mathcal{R}\alpha\}$ zu
3. Falls eines bei 2. erzeugten Schemata einen Kandidatenschlüssel von $\mathcal{R}$ bzgl. Fc enthält, sind wir fertig. Sonst wähle einen Kandidatenschlüssel $\kappa \subseteq \mathcal{R}$ aus und definiere folgendes Schema:
   - $\mathcal{R}\kappa := \kappa$, $F\kappa := \emptyset$
4. Eliminiere diejenigen Schemata $\mathcal{R}\alpha$, die in einem anderen Schema $\mathcal{R}\alpha'$ enthalten sind, d.h., $\mathcal{R}\alpha \subseteq \mathcal{R}\alpha'$

Beispiel

ProfessorenAdr: {[PersNr, Name, Rang, Raum, Ort, Strasse, PLZ, Vorwahl, BLand, EW, Landesregierung]}

FDs

- $\{\text{PersNr}\} \rightarrow \{\text{Name, Rang, Raum, Ort, Strasse, BLand}\}$
- $\{\text{Raum}\} \rightarrow \{\text{PersNr}\}$
- $\{\text{Strasse, BLand, Ort}\} \rightarrow \{\text{PLZ}\}$
- $\{\text{Ort,BLand}\} \rightarrow \{\text{EW, Vorwahl}\}$
- $\{\text{BLand}\} \rightarrow \{\text{Landesregierung}\}$
- $\{\text{PLZ}\} \rightarrow \{\text{BLand, Ort}\}$

Kanonische Überdeckung?

Kandidatenschlüssel?

Beispiel

ProfessorenAdr: $\{[\text{PersNr, Name, Rang, Raum, Ort, Strasse, PLZ, Vorwahl, BLand, EW, Landesregierung}]\}$
FDs

- $\{\text{PersNr}\} \rightarrow \{\text{Name, Rang, Raum, Ort, Strasse, BLand}\}$
- $\{\text{Raum}\} \rightarrow \{\text{PersNr}\}$
- $\{\text{Strasse, BLand, Ort}\} \rightarrow \{\text{PLZ}\}$
- $\{\text{Ort,BLand}\} \rightarrow \{\text{EW, Vorwahl}\}$
- $\{\text{BLand}\} \rightarrow \{\text{Landesregierung}\}$
- $\{\text{PLZ}\} \rightarrow \{\text{BLand, Ort}\}$

Es ist schon eine kanonische Überdeckung.
Kandidatenschlüssel: $\{\text{PersNr}\}$, $\{\text{Raum}\}$.
Ist ProfessorenAdr in 3NF?

Beispiel

ProfessorenAdr: $\{[$PersNr, Name, Rang, Raum, Ort, Strasse, PLZ, Vorwahl, BLand, EW, Landesregierung$]\}$
FDs

- $\{$PersNr$\} \rightarrow \{$Name, Rang, Raum, Ort, Strasse, BLand$\}$
- $\{$Raum$\} \rightarrow \{$PersNr$\}$
- $\{$Strasse, BLand, Ort$\} \rightarrow \{$PLZ$\}$
- $\{$Ort,BLand$\} \rightarrow \{$EW, Vorwahl$\}$
- $\{$BLand$\} \rightarrow \{$Landesregierung$\}$
- $\{$PLZ$\} \rightarrow \{$BLand, Ort$\}$

Kandidatenschlüssel: $\{$PersNr$\}$, $\{$Raum$\}$.
Ist ProfessorenAdr in 3NF? Nein!
FD $\{$Ort,BLand$\} \rightarrow \{$Vorwahl$\}$, wobei

- $\{$Vorwahl$\} \not\subseteq \{$Ort,BLand$\}$
- $\{$Vorwahl$\}$ nicht in einem Kandidatenschlüssel enthalten ist.
- $\{$Ort,BLand$\}$ ist kein Superschlüssel von ProfessorenAddr.

Beispiel

ProfessorenAdr: {[PersNr, Name, Rang, Raum, Ort, Strasse, PLZ, Vorwahl, BLand, EW, Landesregierung]}
FDs

1. {PersNr} → {Name, Rang, Raum, Ort, Strasse, BLand}
2. {Raum} → {PersNr}
3. {Strasse, BLand, Ort} → {PLZ}
4. {Ort,BLand} → {EW, Vorwahl}
5. {BLand} → {Landesregierung}
6. {PLZ} → {BLand, Ort}

Aus FD 1: {Name, Rang, Raum, Ort, Strasse, BLand}. Sie überdeckt auch FD 2.
Aus FD 3: {Strasse, BLand, Ort, PLZ}. Sie überdeckt auch FD 6.
Aus FD 4: {Ort, BLand, EW, Vorwahl}.
Aus FD 5: {BLand, Landesregierung}.
Kandidatenschlüssel: {PersNr}, {Raum} sind schon in einem erzeugten Schema enthalten.

Boyce-Codd Normalform (BCNF)

Schema $\mathcal{R}$ mit FDs F ist in BCNF, wenn für jede FD $(\alpha \rightarrow \beta) \in F$ eine von zwei Bedingungen gilt:

- $\beta \subseteq \alpha$, d.h. die FD ist trivial
- α ist ein Superschlüssel von $\mathcal{R}$

BCNF ist eine Verschärfung der 3NF.
Man kann jedes Schema verlustlos in BCNF-Schemata zerlegen.
Manchmal lässt sich dabei die Abhängigkeiterhaltung aber nicht erzielen.

Beispiel

Städte: $\{[\text{Ort, BLand, Ministerpräsident/in, EW}]\}$
FDs:

1. $\{\text{Ort, BLand}\} \rightarrow \{\text{EW}\}$
2. $\{\text{BLand}\} \rightarrow \{\text{Ministerpräsident/in}\}$
3. $\{\text{Ministerpräsident/in}\} \rightarrow \{\text{BLand}\}$

Kandidatenschlüssel?

Beispiel

Städte: $\{[\text{Ort, BLand, Ministerpräsident/in, EW}]\}$
FDs:

1. $\{\text{Ort, BLand}\} \rightarrow \{\text{EW}\}$
2. $\{\text{BLand}\} \rightarrow \{\text{Ministerpräsident/in}\}$
3. $\{\text{Ministerpräsident/in}\} \rightarrow \{\text{BLand}\}$

Kandidatenschlüssel:

- $\{\text{Ort, BLand}\}$
- $\{\text{Ort, Ministerpräsident/in}\}$

Ist Städte in 3NF?

Beispiel

Städte: $\{[\text{Ort, BLand, Ministerpräsident/in, EW}]\}$
FDs:

1. $\{\text{Ort, BLand}\} \rightarrow \{\text{EW}\}$
2. $\{\text{BLand}\} \rightarrow \{\text{Ministerpräsident/in}\}$
3. $\{\text{Ministerpräsident/in}\} \rightarrow \{\text{BLand}\}$

Schlüsselkandidaten:

- $\{\text{Ort, BLand}\}$
- $\{\text{Ort, Ministerpräsident/in}\}$

Ist Städte in 3NF? **Ja!**

- Die rechten Seiten der FDs 2 und 3 sind in Kandidatenschlüssel enthalten.
- Die linke Seite der FD 1 ist ein Kandidatenschlüssel.

Beispiel

Städte: {[Ort, BLand, Ministerpräsident/in, EW]}
FDs:

1. {Ort, BLand} $\rightarrow$ {EW}
2. {BLand} $\rightarrow$ {Ministerpräsident/in}
3. {Ministerpräsident/in} $\rightarrow$ {BLand}

Schlüsselkandidaten:

- {Ort, BLand}
- {Ort, Ministerpräsident/in}

Ist Städte in BCNF?

Beispiel

Städte: {[Ort, BLand, Ministerpräsident/in, EW]}
FDs:

1. {Ort, BLand} $\rightarrow$ {EW}
2. {BLand} $\rightarrow$ {Ministerpräsident/in}
3. {Ministerpräsident/in} $\rightarrow$ {BLand}

Schlüsselkandidaten:

- {Ort, BLand}
- {Ort, Ministerpräsident/in}

Ist Städte in BCNF? Nein!

- Die linken Seiten der FDs 2 und 3 sind keine Superschlüssel.

Dekompositions-Algorithmus

- Starte mit $Z = \mathcal{R}$, wobei $\mathcal{R}$ das Eingabeschema ist.
- Solange es noch ein Schema $\mathcal{R}_i$ in Z gibt, das nicht in BCNF ist, mache folgendes:
  - Es gibt also eine für $\mathcal{R}_i$ geltende nicht-triviale FD $\alpha \rightarrow \beta$ mit $\alpha \cap \beta = \emptyset$ und $\alpha \not\rightarrow \mathcal{R}_i$
  - Finde eine solche FD
  - Zerlege $\mathcal{R}_i$ in $\mathcal{R}_{i_1} := \alpha \cup \beta$ und $\mathcal{R}_{i_1} := \mathcal{R}_i - \beta$
  - Ersetze in Z $\mathcal{R}_i$ durch $\mathcal{R}_{i_1}$ und $\mathcal{R}_{i_2}$

Beispiel

Städte: $\{[\text{Ort, BLand, Ministerpräsident/in, EW}]\}$
FDs:

1. $\{\text{Ort, BLand}\} \rightarrow \{\text{EW}\}$
2. $\{\text{BLand}\} \rightarrow \{\text{Ministerpräsident/in}\}$
3. $\{\text{Ministerpräsident/in}\} \rightarrow \{\text{BLand}\}$

Aus FD 2: $\{\text{BLand,Ministerpräsident/in}\}$ und $\{\text{Ort, BLand,EW}\}$.
Beide Schemata sind jetzt in BCNF:

- $\{\text{Ort, BLand}\}$ ist ein Superschlüssel von $\{\text{Ort, BLand,EW}\}$
- $\{\text{BLand}\}$ ist ein Superschlüssel von $\{\text{BLand,Ministerpräsident/in}\}$
- $\{\text{Ministerpräsident/in}\}$ ist ein Superschlüssel von $\{\text{BLand,Ministerpräsident/in}\}$

Die Dekomposition ist auch abhängigkeitsbewahrend:

- FD 1 wird $\{\text{Ort, BLand,EW}\}$ zugeordnet.
- FDs 2 und 3 werden $\{\text{BLand,Ministerpräsident/in}\}$ zugeordnet.

Noch ein Beispiel

Schema PLZverzeichnis: {[Strasse, Ort, Bland, PLZ]}

FDs

1. {PLZ} $\rightarrow$ { Ort, BLand}
2. {Strasse, Ort, BLand} $\rightarrow$ {PLZ}

FD 1 verletzt BCNF.
Aus FD 1: {PLZ,Ort,BLand} und {PLZ,Strasse} in BCNF. FD 2 geht aber verloren.