9. Relationale Entwurfstheorie

Vorlesung "Informationssysteme"

Sommersemester 2017

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

Überblick

- "Schlechte" DB-Schemata (Anomalien)
- Funktionale Abhängigkeiten
- Schlüssel
- Zerlegung von Relationen
- Normalformen

"Schlechte" DB-Schemata

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

- Wieso ist dies ein schlechtes Schema?
  - Änderungsanomalien: Sokrates zieht um, von Raum 226 in Raum 338. Was passiert?
  - Einfügeanomalien: Neuer Professor ohne Vorlesungen?
  - Löschanomalien: Letzte Vorlesung eines Profs wird gelöscht. Was passiert?

Funktionale Abhängigkeiten

- Funktionale Abhängigkeit (*functional dependency – FD*) ist eine Bedingung (sem. Konsistenzbedingung) an die möglichen Ausprägungen eines DB-Schemas
  - ergibt sich aus der Semantik der Anwendungswelt!
  - muss in <u>allen möglichen</u> DB-Ausprägungen gültig sein
- Definition Funktionale Abhängigkeit

  Schema $R = \{A, B, C, D\}$, $\alpha \subseteq R$ und $\beta \subseteq R$, Ausprägung R

  $$\alpha \rightarrow \beta$$

  genau dann, wenn $\forall r, s \in R$ gilt: $r.\alpha = s.\alpha \Rightarrow r.\beta = s.\beta$

  D.h. die α-Werte bestimmen die β-Werte funktional (=eindeutig)
  α heißt auch *Determinante* von β

Beispiel

R			
A	B	C	D
a4	b2	c4	d3
a1	b1	c1	d1
a1	b1	c1	d2
a3	b2	c4	d3
a2	b2	c3	d2

Zustand der Relation R erfüllt:

$$\{A\} \rightarrow \{B\}$$
$$\{A\} \rightarrow \{C\}$$
$$\{C, D\} \rightarrow \{B\}$$

aber nicht:

$\{B\} \rightarrow \{C\}$

Notation:

$CD \rightarrow B$

Einhaltung funktionaler Abhängigkeiten

- Alternative Formulierung
  - Die FD $\alpha \rightarrow \beta$ ist in R erfüllt, wenn für jede mögliche Ausprägung c von α gilt:

$$|\pi_{\beta}\ (\sigma_{\alpha=c}(\mathrm{R}))| \leq 1$$

  - "die Menge aller Tupel von R mit $\alpha = c$ (für beliebiges c) projiziert auf β enthält keinen oder genau einen Wert"
  - ansonsten wäre es auch keine "Funktion"
- Daraus folgt ein einfacher Algorithmus:

```
boolean giltFD(Relation R, FD α → β){
    sortiere R nach α Werten
        für jede Gruppe G_i ⊆ R von Tupeln mit Wert α_i aus α:
            falls nicht alle β_i aus β identisch sind:
                return falsch;
        return wahr;
}
```

Schlüssel

- Superschlüssel (Superset – Obermenge)
  - $\alpha \subseteq R$ ist ein Superschlüssel, falls gilt: $\alpha \rightarrow R$
  - trivial: es gilt $R \rightarrow R$
  - Allerdings sind Superschlüssel nicht notwendigerweise minimal!
- Volle funktionale Abhängigkeit:
  - β ist voll funktional abhängig von α (Notation: $\dot{\rightarrow}$) genau dann wenn gilt:
    - $\alpha \rightarrow \beta$
    - α kann nicht mehr verkleinert (=linksreduziert) werden, d.h.
      - $\forall A \in \alpha$ folgt, dass $(\alpha - \{A\}) \rightarrow \beta$ nicht gilt, bzw. alternativ:
      - $\forall A \in \alpha : \neg((\alpha - \{A\}) \rightarrow \beta)$
  - $\alpha \subseteq R$ ist ein Schlüsselkandidat, falls gilt: $\alpha \dot{\rightarrow} R$
- Ein Schlüsselkandidat wird als Primärschlüssel ausgewählt!

Beispiel zur Schlüsselbestimmung

Städte			
Name	BLand	Vorwahl	EW
Frankfurt	Hessen	69	650000
Frankfurt	Brandenburg	335	84000
München	Bayern	89	1200000
Passau	Bayern	851	50000
Saarbrücken	Saarland	681	175000
Kaiserslautern	Rheinland-Pfalz	631	100000
...	...	...	...

- Schlüsselkandidaten von Städte:
  - {Name, BLand}
  - {Name, Vorwahl}

Bestimmung funktionaler Abhängigkeiten

- Relation Professoren: {[PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, BLand, EW, Landesregierung]}
  - {PersNr} → {PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, BLand, EW, Landesregierung}
  - {Ort, BLand} → {EW, Vorwahl}
  - {PLZ} → {BLand, Ort, EW}
  - {BLand, Ort, Straße} → {PLZ}
  - {BLand} → {Landesregierung}
  - {Raum} → {PersNr}
- Außerdem kann abgeleitet werden:
  - {Raum} → {PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, BLand, EW, Landesregierung}
  - {PLZ} → {Landesregierung}

Herleitung weiterer FDs – Armstrong-Axiome

- Aus einer Menge F von FDs sind weitere FDs herleitbar
  - F^+ wird Hülle (closure) von F genannt.
  - F^+ besteht aus allen aus F herleitbaren FDs
- Inferenzregeln, die Amstrong-Axiome, beschreiben die Herleitung:

  $\alpha, \beta, \gamma, \delta$ sind Teilmengen der Attribute aus R
  - Reflexivität: Falls β eine Teilmenge von α ist ($\beta \subseteq \alpha$) dann gilt immer $\alpha \rightarrow \beta$.
    Insbesondere gilt also immer $\alpha \rightarrow \alpha$.
  - Verstärkung: Falls $\alpha \rightarrow \beta$ gilt, dann gilt auch $\alpha\gamma \rightarrow \beta\gamma$. Hierbei stehe z.B. $\alpha\gamma$ für $\alpha \cup \gamma$
  - Transitivität: Falls $\alpha \rightarrow \beta$ und $\beta \rightarrow \gamma$ gilt, dann gilt auch $\alpha \rightarrow \gamma$.
- Die Armstrong-Axiome sind korrekt und vollständig.
  - D.h. Mit Hilfe dieser Axiome können alle gültigen FDs hergeleitet werden.

Weitere Herleitungsregeln

- Nicht notwendig, aber oftmals komfortabel für Herleitungen:
  - Vereinigungsregel:
    Wenn $\alpha \to \beta$ und $\alpha \to \gamma$ gelten, dann auch $\alpha \to \beta\gamma$.
  - Dekompositionsregel:
    Wenn $\alpha \to \beta\gamma$ gilt, dann gelten auch $\alpha \to \beta$ und $\alpha \to \gamma$.
  - Pseudotransitivitätsregel:
    Wenn $\alpha \to \beta$ und $\gamma\beta \to \delta$, dann gilt auch $\alpha\gamma \to \delta$.

Attributhülle

- Die Attributhülle AttrHülle(F, α) einer Attributmenge α bzgl. FDs F ist die Menge von Attributen α^+, für die gilt $\alpha \rightarrow \alpha^+$.
- Gegeben:
  - eine Menge F von FDs
  - eine Menge von Attributen $\alpha \subseteq R$
- Algorithmus:

  $Erg := \alpha;$
  while (Änderungen an Erg) **do**
      for each FD $\beta \rightarrow \gamma$ in F **do**
          if $\beta \subseteq Erg$ **then** $Erg := Erg \cup \gamma;$
  $\alpha^+ := Erg;$

- Es gilt

  $(\alpha \rightarrow \beta) \in F^+ \Leftrightarrow \beta \subseteq \text{AttrHülle}(F, \alpha)$

Beispielanwendung

- Ziel:
  Bestimme, ob κ einen Superschlüssel einer Relation R bzgl. der FDs in F bildet.
- Lösung:
  Durch Aufruf AttrHülle(F, κ) erhalten wir κ^+. Falls $\kappa^+ = R$: κ ist Superschlüssel von R.

Äquivalente FD-Mengen

- Im Allgemeinen: es gibt viele unterschiedliche äquivalente Mengen von funktionalen Abhängigkeiten.

  Zwei Mengen F und G heißen äquivalent (Notation: $F \equiv G$), wenn ihre Hüllen gleich sind, d.h. $F^+ = G^+$.

  Bedeutung: die gleichen Mengen von FDs müssen herleitbar sein.

- Prüfung der Äquivalenz über Attributhüllen (einfacher!)
  - $\forall FD\ (\alpha \to \beta)\ \in\ F$ gilt: $\beta \subseteq$ AttrHülle(G, α) und
  - $\forall FD\ (\alpha \to \beta)\ \in\ G$ gilt: $\beta \subseteq$ AttrHülle(F, α)

  ➔ $F \equiv G$

Kanonische Überdeckung F_c

- Beobachtung: F^+ kann sehr groß sein, viele redundanten Abhängigkeiten
  - in der Praxis unübersichtlich
- Ziel: **kleinstmögliche** Menge F_c (kanonische Überdeckung) finden, so dass immer noch gilt: $F_c^+ = F^+$
- Folgende **Eigenschaften** müssen erfüllt sein:

1. $F_c \equiv F$, d.h., $F_c^+ = F^+$
2. In F_c existieren keine FDs $\alpha \to \beta$, bei denen α oder β überflüssige Attribute enthalten. D.h. es muss gelten:
   a) $\forall A \in \alpha : (F_c - (\alpha \to \beta) \cup ((\alpha - A) \to \beta)) \not\equiv F_c$
   b) $\forall B \in \beta : (F_c - (\alpha \to \beta) \cup (\alpha \to (\beta - B))) \not\equiv F_c$
3. Jede linke Seite einer funktionalen Abhängigkeit in F_c ist einzigartig.
   Standardform, kann durch sukzessive Anwendung der Vereinigungsregel auf FDs der Art $\alpha \to \beta$ und $\alpha \to \gamma$ erzielt werden, so dass die beiden FDs durch $\alpha \to \beta\gamma$ ersetzt werden.

Algorithmus zur Bestimmung von F_c

1. Führe für jede FD $\alpha \rightarrow \beta \in F$ die Linksreduktion durch, also:
   - Überprüfe für alle $A \in \alpha$, ob A überflüssig ist, d.h. ob $\beta \subseteq$ AttrHülle($F, \alpha - A$). Dann kann $\alpha \rightarrow \beta$ durch $(\alpha - A) \rightarrow \beta$ ersetzt werden.
2. Führe für jede (verbliebene) FD die Rechtsreduktion durch:
   - Überprüfe für alle $B \in \beta$, ob
     $$B \in \text{AttrHülle}(F - (\alpha \rightarrow \beta) \cup (\alpha \rightarrow (\beta - B)), \alpha)$$
     gilt. In diesem Fall ist B auf der rechten Seite überflüssig und kann eliminiert werden, d.h. $\alpha \rightarrow \beta$ wird durch $\alpha \rightarrow (\beta - B)$ ersetzt.
3. Entferne die FDs der Form $\alpha \rightarrow \emptyset$, die im 2. Schritt entstanden sind.
4. Fasse mittels Vereinigungsregel FDs der Form $\alpha \rightarrow \beta_1, \ldots,$ $\alpha \rightarrow \beta_n$ zusammen, so dass $\alpha \rightarrow (\beta_1 \cup \cdots \beta_n)$ verbleibt.

Beispiel für Herleitung einer kanonischen Überdeckung

$F = \{A \to B, B \to C, AB \to C\}$

- Schritt 1: Linksreduktion ersetzt $AB \to C$ durch $A \to C$
  - $\{C\} \subseteq$ AttrHülle$(\{A \to B, B \to C, AB \to C\}, A)$?
- Schritt 2: Rechtsreduktion ersetzt $A \to C$ durch $A \to \emptyset$
  - $C \in$ AttrHülle$(\{A \to B, B \to C, A \to \emptyset\}, A)$?
- Schritt 3: eliminiere $A \to \emptyset$
- Schritt 4: keine Zusammenfassungen notwendig

Damit ist $F_c = \{A \to B, B \to C\}$

Normalisierung von Relationen

Um Qualitätsprobleme im ursprünglichen Entwurf zu beheben, wird das bestehende Relationenschema $\mathcal{R}$ in mehrere Relationenschemata $\mathcal{R}_1, \ldots, \mathcal{R}_n$ zerlegt, die dann "besser" sind.

- Die Güte einer Zerlegung wird mit Normalformen beschrieben.
- Normalformen: 1NF, 2NF, 3NF, BCNF, 4NF, ...

▪ Korrektheitskriterien für Zerlegung:

- Verlustlosigkeit: Die in der ursprünglichen Relationenausprägung R des Schemas $\mathcal{R}$ enthaltenen Informationen müssen aus den Ausprägungen $R_1, \ldots, R_n$ der neuen Relationenschemata $\mathcal{R}_1, \ldots, \mathcal{R}_n$ rekonstruierbar sein.
- Abhängigkeitsbewahrung: Alle FDs in $F_{\mathcal{R}}$ sollten in den $F_{\mathcal{R}_1}, F_{\mathcal{R}_2}, \ldots F_{\mathcal{R}_n}$ bewahrt bleiben.

Verlustlosigkeit

- Zerlegung ist gültig, wenn: $R = R_1 \cup R_2$
  - D.h. alle Attribute aus R bleiben in der Zerlegung erhalten
- Wir definieren:
  - $R_1 := \pi_{R_1}(R)$
  - $R_2 := \pi_{R_2}(R)$
- Die Zerlegung von R in R_1 und R_2 ist verlustlos, falls für jede mögliche (gültige) Ausprägung R von R gilt:

$$R = R_1 \bowtie R_2$$

Die Zerlegung muss also durch einen natürlichen Verbund (Join) rekonstruierbar sein.

- D.h. $R_1 \cap R_2 \neq \emptyset$ - "sinnvoller" Schlüssel existiert

Formale Charakterisierung verlustloser Zerlegungen

- Gegeben Zerlegung von R in R_1 und R_2
  - F_R ist die Menge der FDs in R
- Zerlegung ist verlustlos, wenn mindestens eine der folgenden FDs herleitbar ist:
  - $(R_1 \cap R_2) \rightarrow R_1 \in F_R^+$ oder   d.h. Schlüssel bestimmt R_1
  - $(R_1 \cap R_2) \rightarrow R_2 \in F_R^+$ d.h. Schlüssel bestimmt R_2
- Beispiel:
  - Seien α, β und γ paarweise disjunkte Attributmengen
  - $R = \alpha \cup \beta \cup \gamma, R_1 = \alpha \cup \beta, R_2 = \alpha \cup \gamma, R_1 \cap R_2 = \alpha$
  - Dann muss eine der Bedingungen gelten:
    - $\beta \subseteq$ AttrHülle(F_R, α) oder
    - $\gamma \subseteq$ AttrHülle(F_R, α)
- D.h. die gemeinsamen Joinattribute α müssen R_1 oder R_2 bestimmen. Dies ist eine hinreichende aber nicht notwendige Bedingung!

Beispiel für Verlust

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

wird zerlegt in ...

$\pi_{Kneipe,\ Gast}$

Besucht	
Kneipe	**Gast**
Kowalski	Kemper
Kowalski	Eickler
Innsteg	Kemper

$\pi_{Gast,\ Bier}$

Trinkt	
Gast	**Bier**
Kemper	Pils
Eickler	Hefeweizen
Kemper	Hefeweizen

Beispiel für Verlust: Wiederherstellung

$\pi_{Kneipe,\ Gast}$

Besucht	
Kneipe	Gast
Kowalski	Kemper
Kowalski	Eickler
Innsteg	Kemper

$\pi_{Gast,\ Bier}$

Trinkt	
Gast	Bier
Kemper	Pils
Eickler	Hefeweizen
Kemper	Hefeweizen

mit Join verbunden gibt ...

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

Erläuterung des Biertrinker-Beispiels

- Unser Biertrinker-Beispiel war keine verlustlose Zerlegung.
- Es gibt in Biertrinker nämlich nur die folgende nicht-triviale funktionale Abhängigkeit:

$$\{Kneipe, Gast\} \rightarrow \{Bier\}$$

- $R = \{Kneipe, Gast, Bier\}$
- $R_1 = \{Kneipe, Gast\}, R_2 = \{Gast, Bier\}, R_1 \cap R_2 = \{Gast\}$.
- Dann muss eine der Bedingungen gelten:
  - $\{Kneipe\} \subseteq$ AttrHülle$(\{Kneipe, Gast\} \rightarrow \{Bier\}, \{Gast\})$ oder
  - $\{Bier\} \subseteq$ AttrHülle$(\{Kneipe, Gast\} \rightarrow \{Bier\}, \{Gast\})$
- Das ist nicht der Fall!
- Keine der zwei möglichen, die Verlustlosigkeit garantierenden FDs gilt:
  - $\{Gast\} \rightarrow \{Bier\}$
  - $\{Gast\} \rightarrow \{Kneipe\}$

Beispiel für verlustlose Zerlegung

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

wird zerlegt in ...

$\pi_{Vater,\ Kind}$

Väter	
Vater	Kind
Johann	Else
Johann	Theo
Heinz	Cleo

$\pi_{Mutter,\ Kind}$

Mütter	
Mutter	Kind
Martha	Else
Maria	Theo
Martha	Cleo

Erläuterung der Zerlegung der Eltern-Relation

- $R = \{Vater, Mutter, Kind\}$
- $F_R = \{\{Kind\} \rightarrow \{Mutter\}, \{Kind\} \rightarrow \{Vater\}\}$
- $R_1 = \{Vater, Kind\}, R_2 = \{Mutter, Kind\}, R_1 \cap R_2 = \{Kind\}$.
- Dann muss eine der Bedingungen gelten:
  - $\{Vater\} \subseteq$ AttrHülle(F_R , $\{Kind\}$) oder
  - $\{Mutter\} \subseteq$ AttrHülle(F_R , $\{Kind\}$)
- Hier gelten sogar beide Bedingungen.
- Also ist die Zerlegung verlustlos.

Abhängigkeitsbewahrung

- ... ist 2. Korrektheitskriterium für eine Zerlegung
  - R ist zerlegt in R_1 ... R_n
  - Um bei neu eingefügten Daten zu überprüfen, ob alle Abhängigkeiten erfüllt sind, könnte man zur Sicherheit jedesmal den Join $R_1 \bowtie \cdots \bowtie R_n$ berechnen und die Abhängigkeiten überprüfen
  - Das ist allerdings **sehr** teuer!
- Idee: die Abhängigkeiten sollten lokal überprüfbar sein!
  - d.h. Überprüfung kann lokal auf Relationen $R_1, \ldots, R_n$ gemacht werden
  - Dafür muss folgende Bedingung gelten:
    $F_R \equiv (F_{R_1} \cup \cdots \cup F_{R_n})$ bzw. $F_R^+ = (F_{R_1} \cup \cdots \cup F_{R_n})^+$
  - eine solche Zerlegung heisst auch hüllengetreue Dekomposition!

Gegenbeispiel

- **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
    - PLZ Gebiete gehen nicht über Ortsgrenzen und Orte nicht über Bundeslandgrenzen hinweg.
- Daraus ergeben sich die folgenden FDs:
  - $\{PLZ\} \rightarrow \{Ort, Bland\}$
  - $\{Straße, Ort, Bland\} \rightarrow \{PLZ\}$
- Betrachten wir die folgende Zerlegung von PLZverzeichnis:
  - $Straßen$: $\{[PLZ, Straße]\}$
  - $Orte$: $\{[PLZ, Ort, Bland]\}$

Zerlegung der Relation PLZverzeichnis

PLZverzeichnis			
Ort	**Bland**	**Straße**	**PLZ**
Frankfurt	Hessen	Goethestraße	60313
Frankfurt	Hessen	Galgenstraße	60437
Frankfurt	Brandenburg	Goethestraße	15234

$\pi_{\text{PLZ, Straße}}$

Straßen	
PLZ	**Straße**
60313	Goethestraße
60437	Galgenstraße
15234	Goethestraße

$\pi_{\text{Ort, Bland, PLZ}}$

Orte		
Ort	**Bland**	**PLZ**
Frankfurt	Hessen	60313
Frankfurt	Hessen	60437
Frankfurt	Brandenburg	15234

- Diese Zerlegung ist **verlustlos** da $\{PLZ\} \rightarrow \{Ort, Bland\}$ und PLZ das einzige gemeinsame Attribut.
- die FD $\{Straße, Ort, Bland\} \rightarrow \{PLZ\}$ ist im zerlegten Schema nicht mehr enthalten: Also: Zerlegung ist **nicht** abhängigkeitserhaltend.

Zerlegung der Relation PLZverzeichnis (2)

$\pi_{PLZ,\ Straße}$

Straßen	
PLZ	Straße
60313	Goethestraße
60437	Galgenstraße
15234	Goethestraße
15235	**Goethestraße**

$\pi_{Ort,\ Bland,\ Straße}$

Orte		
Ort	Bland	PLZ
Frankfurt	Hessen	60313
Frankfurt	Hessen	60437
Frankfurt	Brandenburg	15234
Frankfurt	**Brandenburg**	**15235**

Einfügen der roten Tupel in Zerlegung ist OK ...

Zerlegung der Relation PLZverzeichnis (3)

- Durch einen Join der Zerlegung erhalten wir dann ...

PLZverzeichnis			
Ort	Bland	Straße	PLZ
Frankfurt	Hessen	Goethestraße	60313
Frankfurt	Hessen	Galgenstraße	60437
Frankfurt	Brandenburg	Goethestraße	15234
Frankfurt	**Brandenburg**	**Goethestraße**	**15235**

- Was fällt auf ?
  - Join erzeugt zusätzliches rotes Tupel, das die FD $\{Straße, Ort, Bland\} \rightarrow \{PLZ\}$ in PLZverzeichnis verletzt!
  - aber: nur durch die Ausführung des Joins ist diese Verletzung der FD aufzudecken.

Zerlegung - Zusammenfassung

- Bisher: **Korrektheit** der Zerlegung
  - Damit eine Zerlegung korrekt ist, muss sie verlustlos und abhängigkeitserhaltend sein.
  - Verlustlos = keine Daten gehen verloren oder werden zusätzlich erzeugt bei einem Join der zerlegten Relationen.
  - abhängigkeitserhaltend = FDs existieren nur lokal aber nicht Relationen-übergreifend
- Jetzt: Bewertung der **Güte** von Relationenschemata
  - Zerlegung von Relationenschemata, um eine höhere Güte zu erreichen
  - Güte = 1NF (NF=Normalform), 2NF, 3NF, BCNF, 4NF, ...

Erste Normalform (1NF)

- Intuition: keine mengenwertigen Attributwerte

  Eine Relation R ist in 1NF, wenn sie keine zusammengesetzten, mengenwertigen oder relationenwertigen Domänen hat.

- Gegenbeispiel:

Eltern		
Vater	Mutter	Kind
Johann	Martha	{Else, Lucie}
Johann	Maria	{Theo, Josef}
Heinz	Martha	{Cleo}

- 1NF:

Eltern		
Vater	Mutter	Kind
Johann	Martha	Else
Johann	Martha	Lucie
Johann	Maria	Theo
Johann	Maria	Josef
Heinz	Martha	Cleo

Exkurs: NF^2 - Relationen

- Non-First Normal-Form-Relationen = geschachtelte Relationen

Eltern			
Vater	Mutter	Kinder	
Johann	Martha	Else	5
		Lucie	3
Johann	Maria	Theo	3
		Josef	1
Heinz	Martha	Cleo	9

- Vorteil: keine unnötige Wiederholung (=Redundanz) von Vater und Mutter
- NF^2 eng verwandt mit hierarchischen Datenmodellen (XML)

Im Folgenden setzen wir immer 1NF voraus

Zweite Normalform (2NF)

- Intuition: nicht mehr als ein Konzept pro Relation modellieren!

  Eine Relation R mit zugehörigen FDs F_R ist in 2NF, falls jedes Nichtschlüssel-Attribut $A \in R$ voll funktional abhängig ist von jedem Schlüsselkandidat der Relation.

- Seien $\kappa_1, \ldots, \kappa_i$ die Schlüsselkandidaten von R
- Sei $A \in R - (\kappa_1 \cup \ \ldots \cup \kappa_i\ )$
  - Ein solches Attribut A wird als nicht-prim (Nichtschlüssel-Attribut) bezeichnet.
  - Gegensatz: Schlüsselattribute bezeichnet man als prim.
- Dann muss für alle Schlüsselkandidaten κ_j $(1 \leq j \leq i)$ gelten:

$$\kappa_j \dot{\rightarrow} A \in F_R^+$$

- Also mit anderen Worten:
  - A ist voll funktional abhängig von jedem κ_j
  - κ_j muss bereits linksreduziert sein!
  - 2NF verhindert partielle Abhängigkeiten

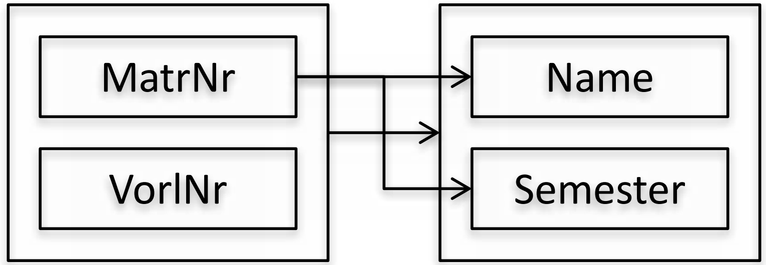
Gegenbeispiel

StudentenBelegung			
MatrNr	VorlNr	Name	Semester
26120	5001	Fichte	10
27550	5001	Schopenhauer	6
27550	4052	Schopenhauer	6
28106	5041	Carnap	3
28106	5052	Carnap	3
28106	5216	Carnap	3
28106	5259	Carnap	3
...	...	...	...

- Schlüsselkandidat $\{MatrNr, VorlNr\}$
  - **prim**: $\{MatrNr, VorlNr\}$
  - **nicht-prim**: $\{Name, Semester\}$
- StudentenBelegung ist **nicht** in 2NF:
  - $\{MatrNr\} \rightarrow \{Name\}$, damit nicht voll funktional abhängig von $\{MatrNr, VorlNr\}$
  - $\{MatrNr\} \rightarrow \{Semester\}$, damit nicht voll funktional abhängig von $\{MatrNr, VorlNr\}$
- Abhängigkeit zu **Teilschlüssel**!

Änderungsanomalien im Gegenbeispiel

StudentenBelegung			
MatrNr	VorlNr	Name	Semester
26120	5001	Fichte	10
27550	5001	Schopenhauer	6
27550	4052	Schopenhauer	6
28106	5041	Carnap	3
28106	5052	Carnap	3
28106	5216	Carnap	3
28106	5259	Carnap	3
...	...	...	...



- **Einfügeanomalie**: Was macht man mit Studenten, die keine Vorlesungen hören?
- **Updateanomalien**: Wenn z.B. Carnap ins vierte Semester kommt, muss man sicherstellen, dass alle vier Tupel geändert werden.
- **Löschanomalie**: Was passiert wenn Fichte seine einzige Vorlesung absagt?
- Zerlegung in zwei Relationen:
  - $hoeren: \{[MatrNr, VorlNr]\}$
  - $Studenten: \{[MatrNr, Name, Semester]\}$
- Beide Relationen sind in **2NF**.

Dritte Normalform (3NF)

- Intuition: Nicht-Schlüssel-Attribut darf kein anderes Nicht-Schlüssel-Attribut bestimmen.

  Ein Relationenschema R ist in 3NF, wenn für jede für R geltende FD der Form $\alpha \rightarrow B$ mit Attribut $B \in R$ mindestens eine von drei Bedingungen gilt:

  1. $B \in \alpha$, d.h. die FD ist trivial
  2. α is Superschlüssel von R
  3. B ist prim

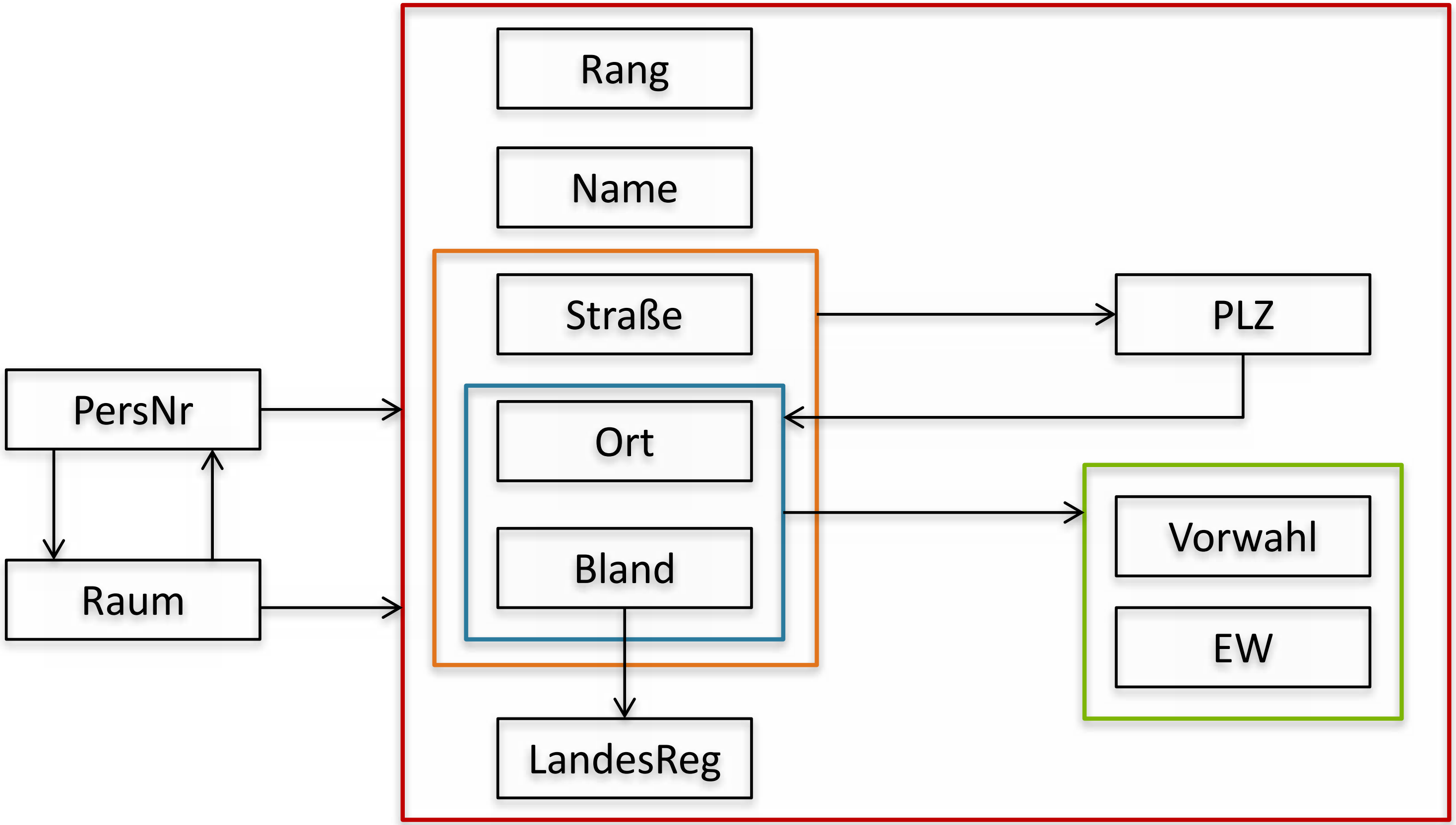
- Eigenschaften:
  - 3NF verhindert partielle und transitive Abhängigkeiten
  - 3NF $\Rightarrow$ 2NF

Gegenbeispiel: Nicht 2NF

StudentenBelegung			
MatrNr	VorlNr	Name	Semester
26120	5001	Fichte	10
27550	5001	Schopenhauer	6
27550	4052	Schopenhauer	6
28106	5041	Carnap	3
28106	5052	Carnap	3
28106	5216	Carnap	3
28106	5259	Carnap	3
…	…	…	…

- Schlüsselkandidat $\{MatrNr, VorlNr\}$
- $\{MatrNr\} \rightarrow \{Name\}$
  - nicht trivial
  - $MatrNr$ kein Superschlüssel
  - $Name$ ist nicht prim
  - Also: nicht in 3NF
- analog für $\{MatrNr\} \rightarrow \{Semester\}$
- Abhängigkeit zu Teilschlüssel durch Superschlüssel-Bedingung in 3NF-Definition abgefragt!

Gegenbeispiel: 2NF aber nicht 3NF



Gegenbeispiel: 2NF aber nicht 3NF (2)

- ProfessorenAdr: $\{[PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, Bland, EW, Landesregierung]\}$
- Schlüsselkandidaten: $PersNr$ und $Raum$
- FDs:
  1. $\{PersNr\} \rightarrow \{Name, Rang, Raum, Ort, Straße, Bland\}$
  2. $\{Raum\} \rightarrow \{PersNr\}$
  3. $\{Straße, Bland, Ort\} \rightarrow \{PLZ\}$
  4. $\{Ort, Bland\} \rightarrow \{EW, Vorwahl\}$
  5. $\{Bland\} \rightarrow \{Landesregierung\}$
  6. $\{PLZ\} \rightarrow \{Bland, Ort\}$

Warum ist diese Relation in 2NF?

Warum ist diese Relation nicht in 3NF?

Gegenbeispiel: 2NF aber nicht 3NF (3)

- $ProfessorenAdr: \{[PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, Bland, EW, Landesregierung]\}$
- Schlüsselkandidaten: $PersNr$ und $Raum$
- FDs:
  1. $\{PersNr\} \rightarrow \{Name, Rang, Raum, Ort, Straße, Bland\}$
  2. $\{Raum\} \rightarrow \{PersNr\}$
  3. $\{Straße, Bland, Ort\} \rightarrow \{PLZ\}$
  4. $\{Ort, Bland\} \rightarrow \{EW, Vorwahl\}$
  5. $\{Bland\} \rightarrow \{Landesregierung\}$
  6. $\{PLZ\} \rightarrow \{Bland, Ort\}$
- FD1 ist OK: $PersNr$ ist Superschlüssel, $Raum$ prim
- FD2 ist OK: $Raum$ ist Superschlüssel, $PersNr$ prim
- FD3-6 nicht OK: weder trivial, noch Superschlüssel, noch prim.

Synthesealgorithmus

- Algorithmus ermittelt zu einem gegebenen Relationenschema R mit funktionalen Abhängigkeiten F eine Zerlegung in $R_1, \dots, R_n$ die alle drei folgenden Kriterien erfüllt:
    1. $R_1, \dots, R_n$ ist eine verlustlose Zerlegung von R.
    2. Die Zerlegung $R_1, \dots, R_n$ ist abhängigkeitserhaltend.
    3. Alle $R_1, \dots, R_n$ sind in 3NF.

Synthesealgorithmus

1. Bestimme die kanonische Überdeckung F_c zu F. Wiederholung:
   i. Linksreduktion
   ii. Rechtsreduktion
   iii. Entfernung von FDs der Form $\alpha \rightarrow \emptyset$
   iv. Zusammenfassung gleicher linker Seiten
2. Für jede funktionale Abhängigkeit $\alpha \rightarrow \beta \in F_c$:
   - Kreiere ein Relationenschema $R_\alpha := \alpha \cup \beta$
   - Ordne R_α die FDs $F_\alpha := \{\alpha' \rightarrow \beta' \in F_c | \alpha' \cup \beta' \subseteq R_\alpha\}$ zu.
3. Falls eines der in (2) erzeugten Schemata einen Schlüsselkandidaten von R bzgl. F_c enthält, sind wir fertig. Sonst wähle einen Schlüsselkandidat $\kappa \subseteq R$ aus und definiere folgendes Schema:
   - $R_\kappa := \kappa$
   - $F_\kappa := \emptyset$
4. Eliminiere diejenigen Schemata R_α , die in einem anderen Relationenschema $R_{\alpha'}$ enthalten sind, d.h., $R_\alpha \subseteq R_{\alpha'}$

Anwendung: Schritt 1

$ProfessorenAdr: \; \{[PersNr, Name, Rang, Raum, Ort, Straße, PLZ, \; Vorwahl, Bland, EW, Landesregierung]\}$

- Schritt 1 (kanonische Überdeckung) enthält die FDs:
  1. $\{PersNr\} \rightarrow \{Name, Rang, Raum, Ort, Straße, Bland\}$
  2. $\{Raum\} \rightarrow \{PersNr\}$
  3. $\{Straße, Bland, Ort\} \rightarrow \{PLZ\}$
  4. $\{Ort, Bland\} \rightarrow \{EW, Vorwahl\}$
  5. $\{Bland\} \rightarrow \{Landesregierung\}$
  6. $\{PLZ\} \rightarrow \{Bland, Ort\}$

Anwendung: Schritt 2

- Schritt 2: Aus den FDs Relationen erzeugen
  1. $\{PersNr\} \rightarrow \{Name, Rang, Raum, Ort, Straße, Bland\}$
     $Professoren: \{[PersNr, Name, Rang, Raum, Ort, Straße, Bland]\}$
  2. $\{Raum\} \rightarrow \{PersNr\}$
     $ProfessorenRäume: \{[Raum, PersNr]\}$
  3. $\{Straße, Bland, Ort\} \rightarrow \{PLZ\}$
     $PLZverzeichnis: \{[Straße, Bland, Ort, PLZ]\}$
  4. $\{Ort, Bland\} \rightarrow \{EW, Vorwahl\}$
     $StädteVerzeichnis: \{[Ort, Bland, EW, Vorwahl]\}$
  5. $\{Bland\} \rightarrow \{Landesregierung\}$
     $Regierung: \{[Bland, Landesregierung]\}$
  6. $\{PLZ\} \rightarrow \{Bland, Ort\}$
     $PLZverzeichnis2: \{[PLZ, Bland, Ort]\}$

Anwendung: Schritt 3

- Schritt 3: Neue Relation falls keine Rel. mit Kand.-Schlüssel:
  1. $\{PersNr\} \rightarrow \{Name, Rang, Raum, Ort, Straße, Bland\}$
     $Professoren: \{[\boldsymbol{PersNr}, Name, Rang, \boldsymbol{Raum}, Ort, Straße, Bland]\}$
  2. $\{Raum\} \rightarrow \{PersNr\}$
     $ProfessorenRäume: \{[\boldsymbol{Raum}, \boldsymbol{PersNr}]\}$
  3. $\{Straße, Bland, Ort\} \rightarrow \{PLZ\}$
     $PLZverzeichnis: \{[Straße, Bland, Ort, PLZ]\}$
  4. $\{Ort, Bland\} \rightarrow \{EW, Vorwahl\}$
     $StädteVerzeichnis: \{[Ort, Bland, EW, Vorwahl]\}$
  5. $\{Bland\} \rightarrow \{Landesregierung\}$
     $Regierung: \{[Bland, Landesregierung]\}$
  6. $\{PLZ\} \rightarrow \{Bland, Ort\}$
     $PLZverzeichnis2: \{[PLZ, Bland, Ort]\}$
- $\boldsymbol{Raum}$ und $\boldsymbol{PersNr}$ sind beide Schlüsselkandidaten.
- Schritt 3 erzeugt für dieses Beispiel keine neue Relation.

Anwendung: Schritt 4

- Schritt 4: Redundante Relationenschemata eliminieren

  1. $Professoren$: $\{[\boldsymbol{PersNr}, Name, Rang, \boldsymbol{Raum}, Ort, Straße, Bland]\}$
  2. $\cancel{ProfessorenRäume\text{: } \{[\boldsymbol{Raum}, \boldsymbol{PersNr}]\}}$ $\subseteq$ $Professoren$
  3. $PLZverzeichnis$: $\{[Straße, Bland, Ort, PLZ]\}$
  4. $StädteVerzeichnis$: $\{[Ort, Bland, EW, Vorwahl]\}$
  5. $Regierung$: $\{[Bland, Landesregierung]\}$
  6. $\cancel{PLZverzeichnis2\text{: } \{[PLZ, Bland, Ort]\}}$ $\subseteq$ $PLZverzeichnis$

Fertig!

Anderes Beispiel für Schritt 3

$StudentenBelegung(\boldsymbol{MatrNr}, \boldsymbol{VorlNr}, Name, Semester)$

- kanonische Überdeckung hierfür:

  $$\{MatrNr\} \rightarrow \{Name, Semester\}$$

- daraus folgt die Relation:

  $$Student(\boldsymbol{MatrNr}, Name, Semester)$$

- Student enthält keinen Schlüsselkandidat.
- deswegen Schritt 3: $hören(\boldsymbol{MatrNr}, \boldsymbol{VorlNr})$
- Ansonsten würde die Information wer welche VL hört wegfallen.

Wiederholung: Dritte Normalform (3NF)

- Intuition: Nicht-Schlüssel-Attribut darf kein anderes Nicht-Schlüssel-Attribut bestimmen.

  Ein Relationenschema R ist in 3NF, wenn für jede für R geltende FD der Form $\alpha \rightarrow B$ mit Attribut $B \in R$ mindestens eine von drei Bedingungen gilt:

  1. $B \in \alpha$, d.h. die FD ist trivial
  2. α is Superschlüssel von R
  3. B ist prim

- Eigenschaften:
  - 3NF verhindert partielle und transitive Abhängigkeiten
  - 3NF $\Rightarrow$ 2NF

Boyce-Codd-Normalform (BCNF)

- Die BCNF ist nochmals eine Verschärfung der 3NF. **Intuition**: Jedes Attribut darf **nur** den gesamten Schlüssel beschreiben und nichts anderes.

  Ein Relationenschema R mit FDs F ist in BCNF, wenn für jede für R geltende funktionale Abhängigkeit der Form $\alpha \rightarrow B$ mit Attribut $B \in R$ mindestens **eine** von zwei Bedingungen gilt:

  1. $B \in \alpha$, d.h. die FD ist trivial
  2. α ist Superschlüssel von R

- Unterschied zu 3NF: 3. Bedingung fällt weg (B ist prim).
- Man kann jede Relation verlustlos in BCNF-Relationen zerlegen
- Aber: manchmal lässt sich dabei die Abhängigkeitserhaltung nicht erzielen!

Städte ist in 3NF, aber nicht in BCNF

$Städte$: $\{[Ort, Bland, Ministerpräsident/in, EW]\}$

- Geltende FDs:
  1. $\{Ort, Bland\} \rightarrow \{EW\}$
  2. $\{Bland\} \rightarrow \{Ministerpräsident/in\}$
  3. $\{Ministerpräsident/in\} \rightarrow \{Bland\}$
- Schlüsselkandidaten:
  - $\{Ort, Bland\}$
  - $\{Ort, Ministerpräsident/in\}$
- Ist in 3NF:
  - linke Seite von FD1 ist Superschlüssel (Bedingung 2)
  - rechten Seiten von FD2 und FD3 sind prim (Bedingung 3)
- Ist nicht in BCNF:
  - linke Seiten von FD2 und FD3 sind aber kein Superschlüssel
  - Wer welches Bundesland regiert wird hier also mehrfach abgespeichert
- D.h.: BCNF verhindert zusätzlich zu 3NF transitive Abhängigkeiten zu Schlüssel-Attributen

Dekomposition

Man kann grundsätzlich jedes Relationenschema R mit funktionalen Abhängigkeiten F so in $R_1, \dots, R_n$ zerlegen, dass gilt

- $R_1, \dots, R_n$ ist eine verlustlose Zerlegung von R
- Alle $R_1, \dots, R_n$ sind in BCNF.

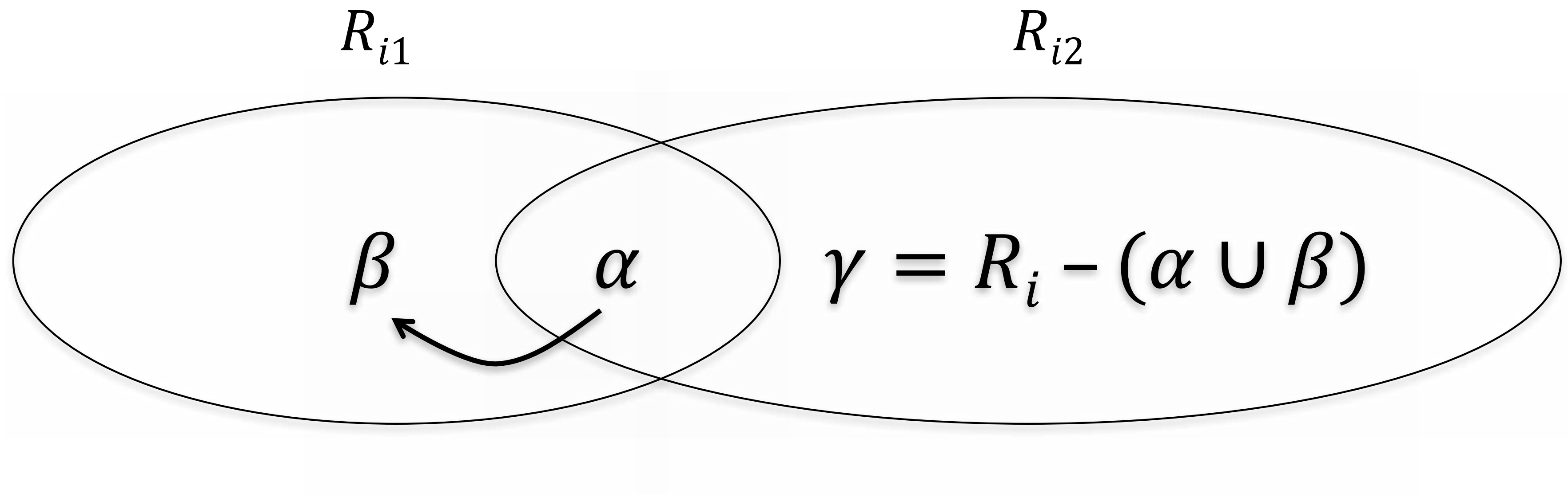
Es kann allerdings nicht immer erreicht werden, dass die Zerlegung $R_1, \dots, R_n$ abhängigkeitserhaltend ist.

Dies ist in der Praxis selten.

Dekompositions-Algorithmus für BCNF

- Starte mit $Z = \{R\}$
- Solange es noch ein Relationenschema $R_i \in Z$ gibt, das nicht in BCNF ist, mache folgendes:
  1. Es gibt also eine für R_i geltende nicht-triviale funktionale Abhängigkeit $\alpha \to \beta$ mit:
     - $\alpha \cap \beta = \emptyset$ (α und β sind disjunkt)
     - $\neg(\alpha \to R_i)$ (α kein Superschlüssel von R_i)
  2. Finde eine solche FD:
     Man sollte sie so wählen, dass β alle von α funktional abhängigen Attribute $B \in (R_i - \alpha)$ enthält, damit der Dekompositionsalgorithmus möglichst schnell terminiert.
  3. Zerlege R_i in $R_{i1} := \alpha \cup \beta$ und $R_{i2} := R_i - \beta$
  4. Entferne R_i aus Z und füge R_{i1} und R_{i2} ein:
     $Z := (Z - R_i) \cup R_{i1} \cup R_{i2}$

Veranschaulichung eines Dekompositionsschrittes

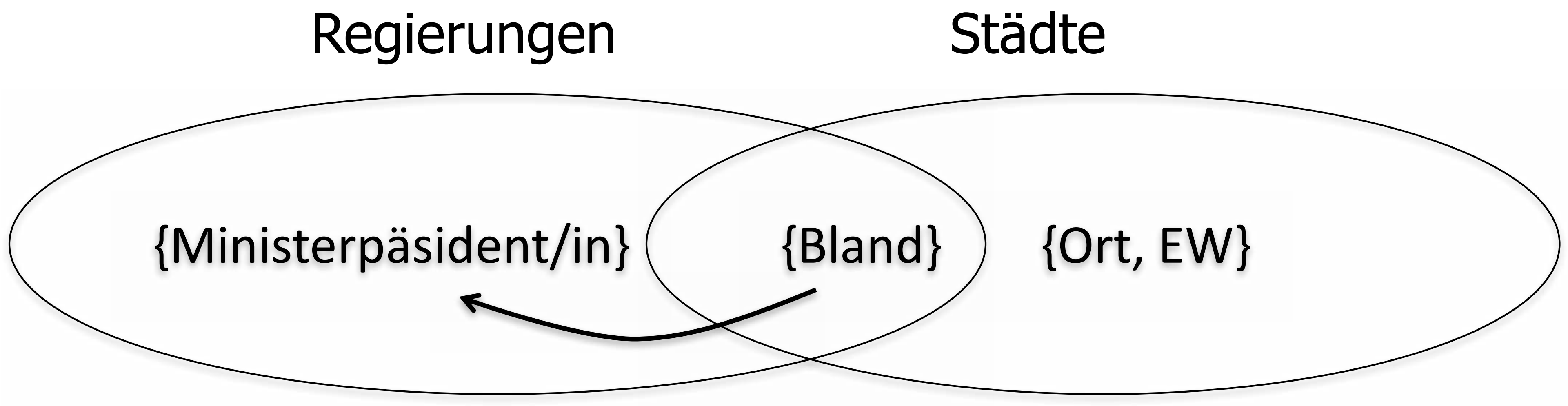


Dekomposition der Relation Städte in BCNF-Relationen

$Städte$: $\{[Ort, Bland, Ministerpräsident/in, EW]\}$

- Geltende FDs:
  1. $\{Ort, Bland\} \rightarrow \{EW\}$
  2. $\{Bland\} \rightarrow \{Ministerpräsident/in\}$
  3. $\{Ministerpräsident/in\} \rightarrow \{Bland\}$
- R_{i1}
  $\boldsymbol{Regierungen}$: $\{[Bland, Ministerpräsident/in]\}$
- R_{i2}
  $\boldsymbol{Städte}$: $\{[Ort, Bland, EW]\}$

Veranschaulichung



- Zerlegung ist verlustlos und auch abhängigkeitserhaltend.
- warum?

Weiteres Beispiel

Dekomposition des $PLZVerzeichnis$ in BCNF-Relationen

- $PLZVerzeichnis$: $\{[Straße, Ort, Bland, PLZ]\}$
  1. $\{Straße, Ort, Bland\} \rightarrow \{PLZ\}$ OK
  2. $\{PLZ\} \rightarrow \{Ort, Bland\}$ verletzt BCNF warum?
- Betrachte nun die Zerlegung:
  - $Orte$: $\{[PLZ, Ort, Bland]\}$
  - $Straßen$: $\{[PLZ, Straße]\}$
- Diese Zerlegung
  - ist verlustlos, aber
  - **nicht abhängigkeitserhaltend, warum?**

Mehrwertige Abhängigkeiten

- Bislang: Funktionale Abhängigkeiten der Form
  - Ausprägung R
  - Seien $\alpha \subseteq R$ und $\beta \subseteq R$
  - $\alpha \rightarrow \beta$ genau dann, wenn $\forall r, s \in R$ gilt:
    $$r.\alpha = s.\alpha \Rightarrow r.\beta = s.\beta$$
  - D.h. die α-Werte bestimmen die β-Werte funktional (=eindeutig)
- Nun: Mehrwertige Abhängigkeiten (multivalued dependencies)
  Notation: $\alpha \twoheadrightarrow \beta$
  Falls einem Attributwert α eine Menge von β-Werten zugeordnet werden.

Genaue Defintion folgt (gleich).

Mehrwertige Abhängigkeiten: Beispiel

Fähigkeiten		
PersNr	Sprache	ProgSprache
3002	griechisch	C
3002	lateinisch	Pascal
3002	griechisch	Pascal
3002	lateinisch	C
3005	deutsch	Ada

- Mehrwertige Abhängigkeiten dieser Relation:
  - $\{PersNr\} \twoheadrightarrow \{Sprache\}$ und
  - $\{PersNr\} \twoheadrightarrow \{ProgSprache\}$
- MVDs führen zu Redundanz und Anomalien

Mehrwertige Anhängigkeiten

R		
A	B	C
a	b	c
a	bb	cc
a	b	cc
a	bb	c

- $A \twoheadrightarrow B$
- $A \twoheadrightarrow C$

Bei zwei Tupeln mit gleichen α-Werten kann man die β-Werte vertauschen und die resultierenden Tupel müssen auch in der Relation sein.

Mehrwertige Anhängigkeiten: Definition 1

	R		
	α $\boxed{A_1, \ldots A_i}$	β $\boxed{A_{i+1}, \ldots A_j}$	γ $\boxed{A_{j+1}, \ldots A_n}$
t_1	$a_1, \ldots a_i$	$a_{i+1}, \ldots a_j$	$a_{j+1}, \ldots a_n$
t_2	$a_1, \ldots a_i$	$b_{i+1}, \ldots b_j$	$b_{j+1}, \ldots b_n$
t_3	$a_1, \ldots a_i$	$a_{i+1}, \ldots a_j$	$b_{j+1}, \ldots b_n$
t_4	$a_1, \ldots a_i$	$b_{i+1}, \ldots b_j$	$a_{j+1}, \ldots a_n$

$\alpha \twoheadrightarrow \beta$ gilt genau dann, wenn für jede Ausprägung von R gilt:

- wenn es zwei Tupel t_1 und t_2 mit gleichen α-Werten gibt, dann muss es auch zwei Tupel t_3 und t_4 geben mit
  - $t_1.\alpha = t_2.\alpha = t_3.\alpha = t_4.\alpha$ (alle α-Werte gleich)
  - $t_3.\beta = t_1.\beta$, $t_4.\beta = t_2.\beta$ (β-Paare gleich)
  - $t_3.\gamma = t_2.\gamma$, $t_4.\gamma = t_1.\gamma$ (γ-Paare vertauscht)

Veranschaulichung: Spezialfall

Veranschaulichung für MVD $\alpha \twoheadrightarrow \beta$:

- Wenn α, β, γ jeweils nur aus einem Attribut A, B, und C bestehen:
- Wenn $\{b_1, \ldots, b_i\}$ und $\{c_1, \ldots, c_j\}$ die B bzw. C -Werte für einen bestimmen A-Wert a sind, dann muss die Relation auch die folgenden $(i * j)$ Tupel enhalten:

$$\{a\} \times \{b_1, \ldots, b_i\} \times \{c_1, \ldots, c_j\}$$

Mehrwertige Abhängigkeiten: Definition 2

- Eine mehrwertige Abhängigkeit (multivalued dependency, MVD) $\alpha \twoheadrightarrow \beta$ besagt, dass einem Attribut α in R eine Menge von β-Werten zugeordnet werden.
- Wenn die MVD $\alpha \twoheadrightarrow \beta$ in R erfüllt ist, dann kann es als Erweiterung zu FDs α, β, c geben mit
$$|\pi_\beta\ (\sigma_{\alpha=c}(R))| > 1$$
- Diese Zuordnung ist unabhängig von den restlichen Attributen in R

Mit anderen Worten:

- α bestimmt nicht nur einen einzelnen Wert (ein singleton)
- genau das ist ja bei einer normalen FD $\alpha \rightarrow \beta$ der Fall!
- sondern eine Menge von Werten
- diese Wertemenge ist unabhängig von den anderen Attributen in $\gamma = R - \alpha - \beta$

Natürlich: Jede FD ist auch eine MVD (aber nicht umgekehrt)

Beispiel

Fähigkeiten		
PersNr	Sprache	ProgSprache
3002	griechisch	C
3002	lateinisch	Pascal
3002	griechisch	Pascal
3002	lateinisch	C
3005	deutsch	Ada

$\pi_{\text{PersNr, Sprache}}$

Sprachen	
PersNr	Sprache
3002	griechisch
3002	lateinisch
3005	deutsch

$\pi_{\text{PersNr, ProgSprache}}$

ProgSprachen	
PersNr	ProgSprache
3002	C
3002	Pascal
3005	Ada

Verlustlose Zerlegung bei MVDs

Hinreichende + notwendige Bedingung:

- Die Zerlegung von R in R_1 und R_2 ist verlustlos genau dann, wenn
  - $R = R_1 \cup R_2$
  - **und** mindestens eine von zwei MVDs gilt:
    - $R_1 \cap R_2 \twoheadrightarrow R_1$ **oder**
    - $R_1 \cap R_2 \twoheadrightarrow R_2$
- Für unser Beispiel gilt:
  - $\{PersNr, Sprache, ProgrSprache\} = \{PersNr, Sprache\} \cup \{PersNr, ProgSprache\}$
  - $\{PersNr\} \twoheadrightarrow \{Sprache\}$
  - $\{PersNr\} \twoheadrightarrow \{ProgSprache\}$
- D.h. es gelten sogar beide MVDs!

MVDs in Paaren

- Es gilt zusätzlich: wenn $\alpha \twoheadrightarrow \beta$, dann gilt immer
  - $\alpha \twoheadrightarrow \gamma$
  - mit $\gamma = R - \alpha - \beta$
- D.h. MVDs treten immer als Paare auf
- Wir könnten MVDs deshalb auch so notieren: $\alpha \twoheadrightarrow \beta | \gamma$

Triviale MVDs ...

- .... sind solche, die von jeder Relationenausprägung R von $\mathcal{R}$ erfüllt werden.
- $\alpha \cup \beta \subseteq \mathcal{R}$
- Eine MVD $\alpha \twoheadrightarrow \beta$ ist trivial genau dann, wenn
  1. $\beta \subseteq \alpha$ oder
  2. $\beta = \mathcal{R} - \alpha$
- Nur die Bedigung 1 galt auch für normale FDs.
- Beispiel für Bedingung 2:
  - $\mathcal{R} = \{PersNr, Sprache\}$
  - $\alpha = \{PersNr\}$
  - $\beta = \{Sprache\}$
  - $\mathcal{R} - \alpha = \{PersNr, Sprache\} - \{PersNr\} = \{Sprache\} = \beta \Rightarrow$ MVD ist trivial!

Vierte Normalform (4NF)

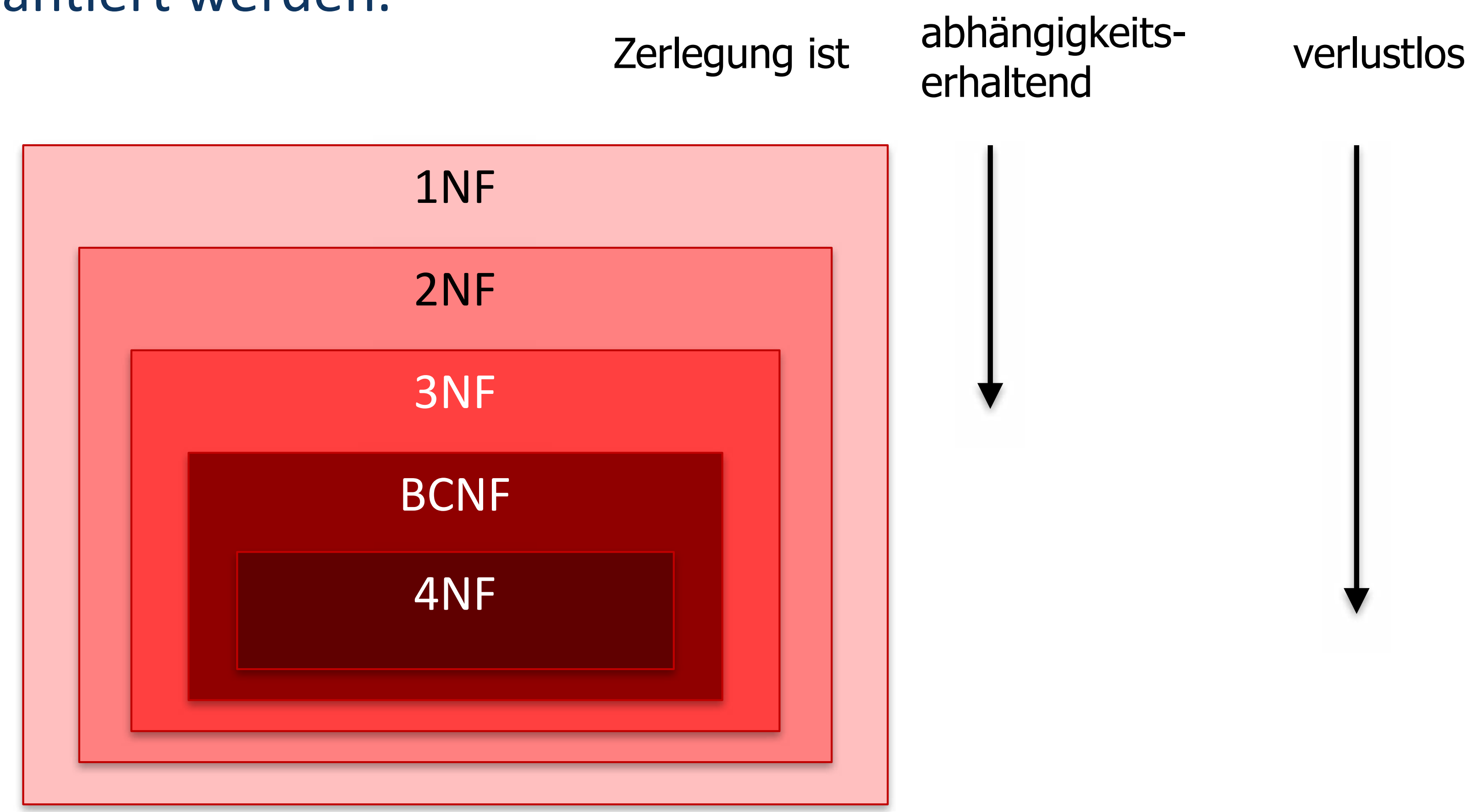
- Eine Relation R ist in 4NF, wenn für jede MVD $\alpha \twoheadrightarrow \beta$ eine der folgenden Bedingungen gilt:
  - Die MVD ist trivial **oder**
  - α ist Superschlüssel von R
- D.h. 4NF ist sehr ähnlich zu BCNF!
- Unterschied:
  - MVDs statt FDs
  - Definition von "trivial" wurde erweitert.
- 4NF erfüllt $\Rightarrow$ BCNF erfüllt, da jede FD eine MVD ist.

Dekomposition in 4NF

- Starte mit der Menge $Z := \{R\}$
- Solange es noch ein Relationenschema R_i in Z gibt, das nicht in 4NF ist, mache folgendes:
  - Es gibt also eine für R_i geltende nicht-triviale MVD $(\alpha \twoheadrightarrow \beta)$, für die gilt:
    - $\alpha \cap \beta = \emptyset$
    - $\neg(\alpha \rightarrow R_i)$
  - Finde eine solche MVD
  - Zerlege R_i in $R_{i1} := \alpha \cup \beta$ und $R_{i2} := R_i - \beta$
  - Entferne R_i aus Z und füge R_{i1} und R_{i2} ein, also $Z := (Z - R_i) \cup \{R_{i1}\} \cup \{R_{i2}\}$

Alle Normalformen auf einen Blick

- Die Verlustlosigkeit ist für alle Zerlegungsalgorithmen in alle Normalformen garantiert.
- Die Abhängigkeitserhaltung kann nur bis zur dritten Normalform garantiert werden.



Zusammenfassung Entwurfstheorie

- Ziel: Bewertung der Güte einer Relation; Vermeidung von Anomalien durch Zerlegung in “bessere” Relationen
- Betrachtung von funktionalen Abhängigkeiten zur Identifikation von Redundanz
- Korrektheit von Zerlegung definiert als verlustlos und abhängigkeitsbewahrend.
- 1NF, 2NF, 3NF, BCNF und 4NF betrachtet.
- Algorithmen zur Zerlegung/Synthese