Datenbanksysteme
7 Relationale Entwurfstheorie und Normalformen

Prof. Dr. Gregor Grambow

Hochschule Aalen
Fakultät Elektronik und Informatik

Überblick

Inhalt

- Funktionale Abhängigkeiten
- Transitive Hülle, Kanonische Überdeckung
- Anomalien
- Normalformen
- Synthesealgorithmus
- Dekompositionsalgorithmus

Ziele

- Theoretische Grundlagen der relationalen Entwurfstheorie verstehen
- Verstehen wofür man Normalformen braucht
- Relationen in die verschiedenen Normalformen bringen können

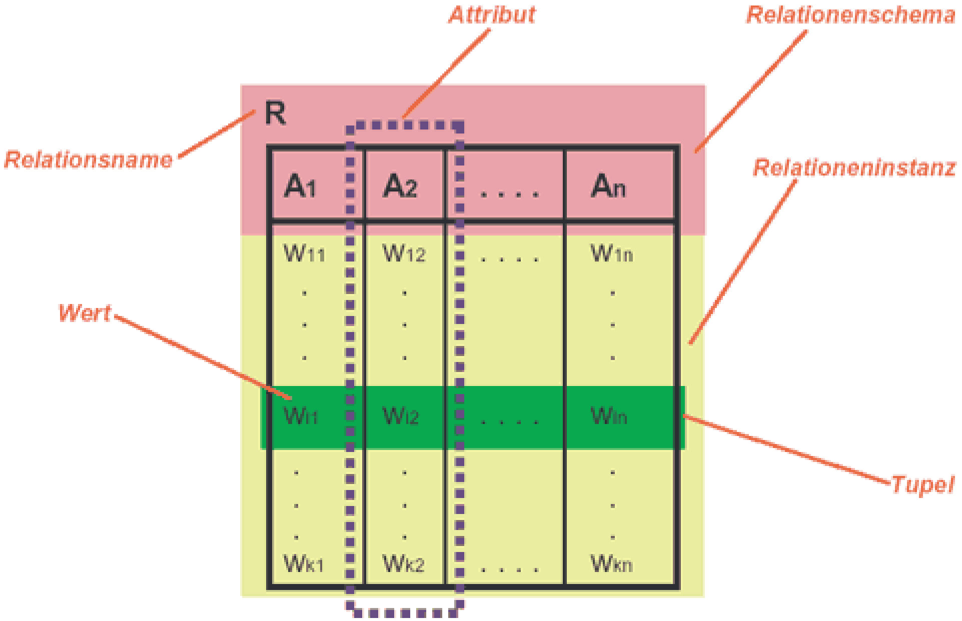
Foliensatz basiert teilweise auf [Kemper, Eickler: Datenbanksysteme – Eine Einführung. Oldenbourg, 2015]

Ziele der relationalen Entwurfstheorie

- Bewertung der Qualität eines Relationenschemas
  - Redundanz
  - Einhaltung von Konsistenzbedingungen
    - Funktionaler Abhängigkeiten
- Normalformen als Gütekriterium
- Ggfls. Verbesserung eines Relationenschemas
  - Durch den Synthesealgorithmus
  - Durch Dekomposition

Zur Notation…

- (Relationen)Schema $\mathcal{R}$ = {A, B, C, D, … , H}
- Attribute: A, B, C, …
- Attributmengen: α, β, ...
- Ausprägung/Relation/Instanz: R (aktueller Inhalt einer Relation)
- Tupel: r, s, t, …
- Projektion: $r.\alpha$, $s.\alpha$, $r.\beta$, $s.\beta$, ...



[http://www.gitta.info/LogicModelin/de/html/RelatiModels_RelatConcept.html]

Funktionale Abhängigkeiten (FDs)

- *„Wenn man von dem Wert eines Attributs auf den Wert eines anderen schließen kann“*
- Schema
  - $\mathcal{R} = \{A, B, C, D\}$
- Ausprägung R

- Seien $\alpha \subseteq \mathcal{R}$, $\beta \subseteq \mathcal{R}$
- $\alpha \rightarrow \beta$ genau dann wenn $\forall r, s \in R$ mit $r.\alpha = s.\alpha \Rightarrow r.\beta = s.\beta$

Beispiel

Stammbaum				
Kind	Vater	Mutter	Opa	Oma
Sofie	Alfons	Sabine	Lothar	Linde
Sofie	Alfons	Sabine	Hubert	Lisa
Niklas	Alfons	Sabine	Lothar	Linde
Niklas	Alfons	Sabine	Hubert	Lisa
...	...	...	Lothar	Martha
...	...	...	...	...

Beispiel

Stammbaum				
Kind	Vater	Mutter	Opa	Oma
Sofie	Alfons	Sabine	Lothar	Linde
Sofie	Alfons	Sabine	Hubert	Lisa
Niklas	Alfons	Sabine	Lothar	Linde
Niklas	Alfons	Sabine	Hubert	Lisa
...	...	...	Lothar	Martha
...	...	...	...	...

- Kind → Vater,Mutter
- Kind,Opa → Oma
- Kind,Oma → Opa

Weiteres Beispiel

- Sind folgende FDs auf der Relation gültig?

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

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

Weiteres Beispiel

- Sind folgende FDs auf der Relation gültig?

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

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

Weiteres Beispiel

- Sind folgende FDs auf der Relation gültig?

- {A} → {B} ja
- {C, D} → {B} ja
- {B} → {C}
- {A, B} → {C}
- {B, C} → {A}
- {B} → {A}

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

Weiteres Beispiel

- Sind folgende FDs auf der Relation gültig?

- $\{A\} \rightarrow \{B\}$ ja
- $\{C, D\} \rightarrow \{B\}$ ja
- $\{B\} \rightarrow \{C\}$ nein
- $\{A, B\} \rightarrow \{C\}$
- $\{B, C\} \rightarrow \{A\}$
- $\{B\} \rightarrow \{A\}$

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

Weiteres Beispiel

- Sind folgende FDs auf der Relation gültig?

- $\{A\} \rightarrow \{B\}$ ja
- $\{C, D\} \rightarrow \{B\}$ ja
- $\{B\} \rightarrow \{C\}$ nein
- $\{A, B\} \rightarrow \{C\}$ ja
- $\{B, C\} \rightarrow \{A\}$
- $\{B\} \rightarrow \{A\}$

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

Weiteres Beispiel

- Sind folgende FDs auf der Relation gültig?

- $\{A\} \rightarrow \{B\}$ ja
- $\{C, D\} \rightarrow \{B\}$ ja
- $\{B\} \rightarrow \{C\}$ nein
- $\{A, B\} \rightarrow \{C\}$ ja
- $\{B, C\} \rightarrow \{A\}$ nein
- $\{B\} \rightarrow \{A\}$

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

Weiteres Beispiel

- Sind folgende FDs auf der Relation gültig?

- $\{A\} \rightarrow \{B\}$ ja
- $\{C, D\} \rightarrow \{B\}$ ja
- $\{B\} \rightarrow \{C\}$ nein
- $\{A, B\} \rightarrow \{C\}$ ja
- $\{B, C\} \rightarrow \{A\}$ nein
- $\{B\} \rightarrow \{A\}$ nein

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

Arten von FDs

- Funktionale Abhängigkeit: $\alpha \rightarrow \beta$ gilt wenn man eindeutig von α auf β schließen kann.
- Volle funktionale Abhängigkeit: $\alpha \rightarrow \beta$ gilt als volle funktionale Abhängigkeit, wenn jedes Element aus β von der kompletten Menge α, nicht von einer echten Teilmenge von α, funktional abhängig ist.
- Transitive Abhängigkeit: Eine transitive Abhängigkeit zwischen α und γ gilt wenn $\alpha \rightarrow \beta$ und $\beta \rightarrow \gamma$ gilt.

Bestimmung von FDs

- Beispiel
- Gegeben: Information über Professoren anhand folgender Attribute.

- Professoren: {[PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, Bland, EW, Landesregierung]}

- Frage: Welche FDs lassen sich aufgrund der Eigenschaften bzw. Semantik des zu modellierenden Ausschnitts der Realwelt finden?

Bestimmung von FDs

- Beispiel
- Professoren: {[PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, Bland, EW, Landesregierung]}
- Frage: Welche FDs lassen sich aufgrund der Eigenschaften bzw. Semantik des zu modellierenden Ausschnitts der Realwelt finden?

- PersNr ist ein Kandidatenschlüssel:

  {PersNr} → {PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, BLand, EW, Landesregierung}

- Ortsnamen sind innerhalb eines Bundeslandes eindeutig:
- {Ort, Bland} → {EW, Vorwahl}

Bestimmung von FDs

- Beispiel
- Professoren: {[PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, Bland, EW, Landesregierung]}

- Die Postleitzahl identifiziert einen Ort, das Bundesland und die Einwohnerzahl:

  {PLZ} → {BLand, Ort, EW}
- Die Postleitzahl ändert sich innerhalb der Straße eines Ortes nicht:

  {BLand, Ort, Straße} → {PLZ}
- Landesregierung speichert die Partei des Ministerpräsidenten:

  {Bland} → {Landesregierung}
- In einem Raum kann nur ein Professor sitzen:

  {Raum} → {PersNr}

Bestimmung von FDs

- <u>Beispiel - Zusammenfassung</u>
- 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}

Bestimmung von FDs

- <u>Beispiel – Abgeleitete FDs</u>

- Gegeben: Menge F von FDs:
  - Raum → PersNr, PersNr → Name

- Frage 1: Welche weiteren FDs können daraus abgeleitet werden?

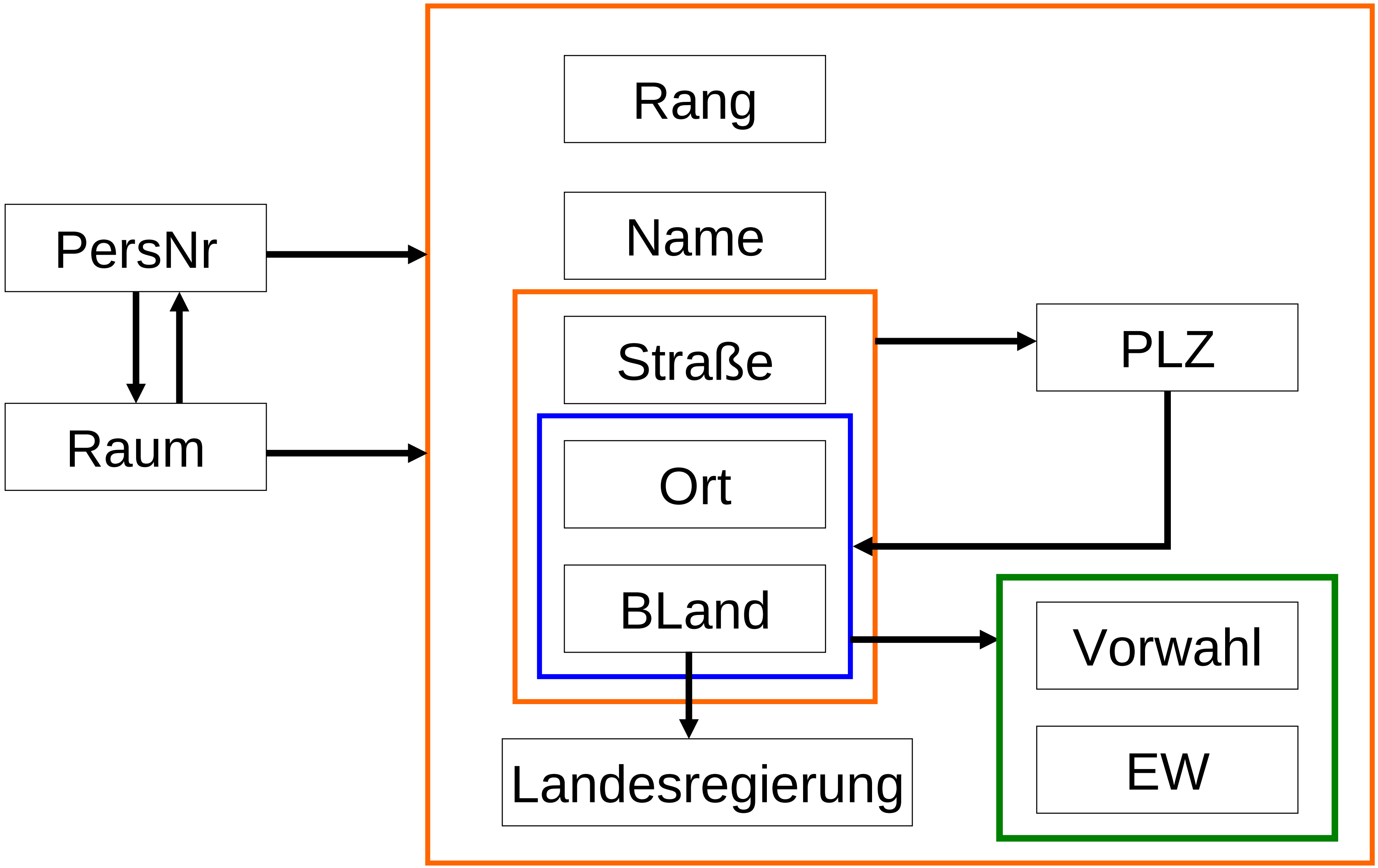
- Frage 2: Wenn zusätzlich eine Menge α an Attributen gegeben ist, welche Attribute kann man aus α aufgrund von F herleiten?

Bestimmung von FDs

- Beispiel Professoren incl. Abgeleiteter FDs
- 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}

- Zusätzliche Abhängigkeiten, die aus obigen abgeleitet werden können:
  - {Raum} → {PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, Bland, EW, Landesregierung}
  - {PLZ} → {Landesregierung}

Graphische Darstellung der funktionalen Abhängigkeiten



Hülle einer Attributmenge

- Die Menge α+ der Attribute, die aus einer Menge von Attributen α aufgrund einer Menge von FDs F hergeleitet werden können, nennt man die Hülle der Attributmenge α.

Bestimmung der Hülle einer Attributmenge

- Eingabe: eine Menge F von FDs und eine Menge von Attributen α.

- Ausgabe: die vollständige Menge von Attributen α+, für die gilt $\alpha \rightarrow \alpha$+.

- AttrHülle(F,α)
  - Erg := α
  - **While** (Änderungen an Erg) **do**
    - **Foreach** FD $\beta \rightarrow \gamma$ **in** F **do**
      - **If** $\beta \subseteq$ Erg **then** Erg := Erg $\cup$ γ
  - Ausgabe α+ = Erg

Hülle einer Attributmenge

- <u>Beispiel</u>
- Sei F = {RS → T, U → VX, RX → W, T → RU}

- AttrHülle(F, {T}) :
- {T, R, U, V, X, W}
- AttrHülle(F, {RS}) :
- {R, S, T, U, V, W, X}

Schlüssel

- $\alpha \subseteq \mathcal{R}$ ist ein Super-Schlüssel, falls folgendes gilt:
  - $\alpha \rightarrow \mathcal{R}$
- β ist voll funktional abhängig von α genau dann wenn gilt
  - $\alpha \rightarrow \beta$ und
  - α kann nicht mehr verkleinert werden, d.h.
    - $\forall A \in \alpha$ folgt, dass $(\alpha - \{A\}) \rightarrow \beta$ nicht gilt, oder kürzer
    - $\forall A \in \alpha: \neg((\alpha - \{A\}) \rightarrow \beta)$
- Notation für volle funktionale Abhängigkeit: $\alpha \stackrel{\bullet}{\rightarrow} \beta$
- $\alpha \subseteq \mathcal{R}$ ist ein Kandidaten-Schlüssel (oder schlicht Schlüssel), falls folgendes gilt:
  - $\alpha \stackrel{\bullet}{\rightarrow} \mathcal{R}$
- Ein Schlüssel ist minimal
- Ein Superschlüssel ist die Obermenge eines Schlüssels

Schlüsselbestimmung

Städte			
Name	BLand	Vorwahl	EW
Frankfurt	Hessen	069	650000
Frankfurt	Brandenburg	0335	84000
München	Bayern	089	1200000
Passau	Bayern	0851	50000
...	...	...	...

- Kandidaten-schlüssel von *Städte*:
  - {Name,BLand}
  - {Name,Vorwahl}

- Beachte, dass 2 kleinere Städte dieselbe Vorwahl haben können

Schlüsselbestimmung mittels AttrHülle

- Suche nach Kandidatenschlüsseln einer Relation R aufgrund der vorhandenen FDs.
- Eine Attributmenge α ist (Kandidaten-)Schlüssel wenn $\alpha \xrightarrow{\bullet} \mathcal{R}$
  - Also wenn AttrHülle(F, α) = $\mathcal{R}$
  - Und die Minimalität erfüllt ist also für jedes Attribut A aus α gilt: AttrHülle(F, $\alpha - \{A\}$) ≠ $\mathcal{R}$

Schlüsselbestimmung mittels AttrHülle

- Beispiel: $\mathcal{R}$ = {ABCDEF}, F = {C → BDAE}
- Zeitraubende Vorgehensweise: Durchprobieren aller einelementigen, aller zweielementigen, aller dreielementigen Schlüsselkandidaten mittels AttrHülle
- Alternativ: Verwendung der folgenden Heuristik: Alle Attribute, die nicht auf der rechten Seite vorkommen, können mittels AttrHülle nicht hergeleitet werden und müssen daher im Schlüssel enthalten sein.
- Allgemeinere Regel:
  - Kommen Attribute ausschließlich auf der linken Seite (oder gar nicht) vor, müssen sie im Schlüsselkandidaten enthalten sein.
  - Kommen Attribute auf beiden Seiten vor, können sie in den Schlüsselkandidaten auftauchen.
  - Kommen Attribute auf der rechten Seite vor, dürfen sie nicht Teil der Schlüsselkandidaten sein.
- Hier: C und F kommen rechts nicht vor, daher folgender Versuch: AttrHülle({C → BDAE}, {CF}) = {C, F, B, D, A, E}
- Also ist CF Schlüssel von $\mathcal{R}$

Schlüsselbestimmung mittels AttrHülle

- Weiteres Beispiel: $\mathcal{R}$ = {ABCDEF},
  F = {C → BD, D → AE, E → CF, F → E}
- Heuristik hier nicht zielführend, da alle Attribute hergeleitet werden können. Also testen wir alle Attribute die auf der linken Seite vorkommen.
- C: AttrHülle({F}, {C}) = {C, B, D, A, E, F}
  - C ist Schlüssel von $\mathcal{R}$
  - Achtung: C kann aus E hergeleitet werden also:
- E: AttrHülle({F}, {E}) = {E, C, B, D, A, F}
  - E ist Schlüssel von $\mathcal{R}$
  - Achtung: E kann aus D hergeleitet werden also:
- D: AttrHülle({F}, {D}) = {D, A, E, C, F, B}
  - D ist Schlüssel von $\mathcal{R}$
  - Achtung: D kann aus F hergeleitet werden also:
- F: AttrHülle({F}, {F}) = {F, E, C, B, D, A}
  - F ist Schlüssel von $\mathcal{R}$

Hülle von FDs

- Problem: F sei Menge von FDs. Welche weiteren FDs können daraus abgeleitet werden?

- Beispiel: {Raum → PersNr, PersNr → Name}
- Abgeleitet: {Raum} → {PersNr, Name}

- Die Menge aller aus F ableitbaren FDs wird Hülle F+ von F genannt.

Herleitung funktionaler Abhängigkeiten: Armstrong-Axiome

- Reflexivität
  - Falls β eine Teilmenge von α ist ($\beta \subseteq \alpha$) dann gilt immer $\alpha \rightarrow \beta$. Insbesondere gilt immer $\alpha \rightarrow \alpha$.
- Verstärkung
  - Falls $\alpha \rightarrow \beta$ gilt, dann gilt auch $\alpha\gamma \rightarrow \beta\gamma$.
- Transitivität
  - Falls $\alpha \rightarrow \beta$ und $\beta \rightarrow \gamma$ gilt, dann gilt auch $\alpha \rightarrow \gamma$.

- Diese drei Axiome sind vollständig und korrekt. Zusätzliche Axiome erleichtern die Herleitung:
  - Vereinigungsregel:
    - Wenn $\alpha \rightarrow \beta$ und $\alpha \rightarrow \gamma$ gelten, dann gilt auch $\alpha \rightarrow \beta\gamma$
  - Dekompositionsregel:
    - Wenn $\alpha \rightarrow \beta\gamma$ gilt, dann gelten auch $\alpha \rightarrow \beta$ und $\alpha \rightarrow \gamma$
  - Pseudotransitivitätsregel:
    - Wenn $\alpha \rightarrow \beta$ und $\gamma\beta \rightarrow \delta$, dann gilt auch $\alpha\gamma \rightarrow \delta$

Kanonische Überdeckung

- Es kann zu einer Menge funktionaler Abhängigkeiten viele verschiedene äquivalente Mengen funktionaler Abhängigkeiten geben. Zwei Mengen funktionaler Abhängigkeiten F und G heißen genau dann äquivalent
  $F \equiv G$,
  wenn ihre Attributhüllen gleich sind
  $F+ = G+$.
  Sind F und G äquivalent, so heißt F *Überdeckung* von G und umgekehrt.
- Es gibt stets eine eindeutige Attributhülle F+ zu einer gegebenen Menge F von funktionalen Abhängigkeiten. Diese Menge beinhaltet in der Regel viele funktionale Abhängigkeiten.
  - Wirkt sich bei einer späteren Implementierung des Schemas in einer relationalen Datenbank negativ aus.
  - Bei jeder Änderungsoperation im Rahmen einer Konsistenzprüfung muss die Einhaltung sämtlicher spezifizierter funktionaler Abhängigkeiten überprüft werden.

Kanonische Überdeckung

- Beim Entwurfsprozess relationaler Schemata ist man an der kleinstmöglichen Menge der äquivalenten funktionalen Abhängigkeiten interessiert.
- Diese nennt man *kanonische Überdeckung* der gegebenen Menge funktionaler Abhängigkeiten.
- Eine *kanonische Überdeckung* beschreibt die kleinste gültige Menge von funktionalen Abhängigkeiten für ein bestimmtes relationales Schema.
- Die Ableitung einer solchen *kanonischen Überdeckung* gewährleistet ein redundanzfreies relationales Schema.
- Die Kanonische Überdeckung ist nicht eindeutig.

Kanonische Überdeckung

- Fc heißt kanonische Überdeckung von F, wenn die folgenden drei Kriterien erfüllt sind:
  1. $Fc \equiv F$, d.h. Fc+ = F+
  2. In Fc existieren keine FDs , die überflüssige Attribute enthalten. D.h. es muss folgendes gelten:
     - $\forall A \in \alpha$: $(Fc - (\alpha \rightarrow \beta) \cup ((\alpha - \{A\}) \rightarrow \beta)) \equiv Fc$
     - $\forall B \in \beta$: $(Fc - (\alpha \rightarrow \beta) \cup (\alpha \rightarrow (\beta - \{B\}))) \equiv Fc$
  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. 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, α - A)

     gilt. Falls dies der Fall ist, ersetze $\alpha \rightarrow \beta$ durch $(\alpha - A) \rightarrow \beta$.
2. Führe für jede (verbliebene) FD die Rechtsreduktion durch, also:

   Ist B bzw. $\alpha \rightarrow B$ überflüssig?

   Vorher Dekomposition durchführen!
   - Überprüfe für alle $B \in \beta$, ob
     - $B \in$ AttrHülle(F – $(\alpha \rightarrow \beta) \cup (\alpha \rightarrow (\beta - B))$, α)

       gilt. Falls dies der Fall ist, ist B auf der rechten Seite überflüssig und kann eliminiert werden, d.h. ersetze $\alpha \rightarrow \beta$ durch $\alpha \rightarrow (\beta - B)$.
3. Entferne die FDs der Form $\alpha \rightarrow \emptyset$, die im 2. Schritt möglicherweise entstanden sind.
4. Fasse mittels der Vereinigungsregel FDs der Form $\alpha \rightarrow \beta 1$, ..., $\alpha \rightarrow \beta n$ zusammen, so dass $\alpha \rightarrow (\beta 1 \cup ... \cup \beta n)$ verbleibt.

Beispiel 1 Berechnung Kanonische Überdeckung

- $F = \{A \rightarrow B, B \rightarrow C, AB \rightarrow C\}$
- Linksreduktion:
  - $A \rightarrow B$: Bereits reduziert
  - $B \rightarrow C$: Bereits reduziert
  - $AB \rightarrow C$: $C \in$ AttrHülle(F, A)
  - Änderung der FD: $F = \{A \rightarrow B, B \rightarrow C, A \rightarrow C\}$
- Rechtsreduktion:
  - $A \rightarrow B$: $B \in$ AttrHülle(F-($A \rightarrow B$), A)? NEIN!
  - $B \rightarrow C$: $C \in$ AttrHülle(F-($B \rightarrow C$), B)? NEIN!
  - $A \rightarrow C$: $C \in$ AttrHülle(F-($A \rightarrow C$), A)? JA!
  - Änderung der FD: $F = \{A \rightarrow B, B \rightarrow C\}$
- Vereinigungsregel nicht anwendbar
- Finale FD: $F = \{A \rightarrow B, B \rightarrow C\}$

Bei FDs auf deren rechter Seite mehr als ein Attribut steht sollte man vorher die Dekompositions-regel anwenden!

Hier nicht nötig!

Beispiel 2 Berechnung Kanonische Überdeckung

- F = {A → BD, AC → E, CD → E, E → A, D → C}
- Dekomposition:
  F = {A → B, A → D, AC → E, CD → E, E → A, D → C}
- Linksreduktion:
  - A → B: ok
  - A → D: ok
  - E → A: ok
  - D → C: ok
  - AC → E: E ∈ AttrHülle(F, A) ja
    Änderung der FD: F = {A → B, A → D, A → E, CD → E, E → A, D → C}
  - CD → E: E ∈ AttrHülle(F, C) nein
    E ∈ AttrHülle(F, D) ja
  - Änderung der FD: F = {A → B, A → D, A → E, D → E, E → A, D → C}

Beispiel 2 Berechnung Kanonische Überdeckung

- F = {A → B, A → D, A → E, D → E, E → A, D → C}
- Rechtsreduktion:
  - A → B: B ∈ AttrHülle(F-{A → B}, A) nein
  - A → D: D ∈ AttrHülle(F-{A → D}, A) nein
  - A → E: E ∈ AttrHülle(F-{A → E}, A) ja

    F = {A → B, A → D, D → E, E → A, D → C}
  - D → E: E ∈ AttrHülle(F-{D → E}, D) nein
  - E → A: A ∈ AttrHülle(F-{E → A}, E) nein
  - D → C: C ∈ AttrHülle(F-{D → C}, D) nein
- Zusammenfassen

  F = {A → BD, D → EC, E → A}

Warum das alles?

„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 R. 338. Was passiert?
- Einfüge-Anomalien
- Neue/r Prof ohne Vorlesungen?
- Löschanomalien
- Letzte Vorlesung einer/s Profs wird gelöscht? Was passiert?

Zerlegung (Dekomposition) von Relationen

- Anomalien beruhen auf der Tatsache, dass nicht zusammenpassende Informationen zusammen gespeichert wurden.
- Intuitiv:
  - alle Informationen zu Professoren werden in einer Relation gespeichert,
  - alle Informationen zu Vorlesungen werden in einer Relation gespeichert,
  - die Information über die Verknüpfung beider Relationen wird in einer Relation gespeichert.
- Lösung: Zerlegung des Schemas in Teilschemata
- Frage: was ist eine „sinnvolle“ Zerlegung?

Zerlegung (Dekomposition) von Relationen

- Es gibt zwei Korrektheitskriterien für die Zerlegung von Relationenschemata:

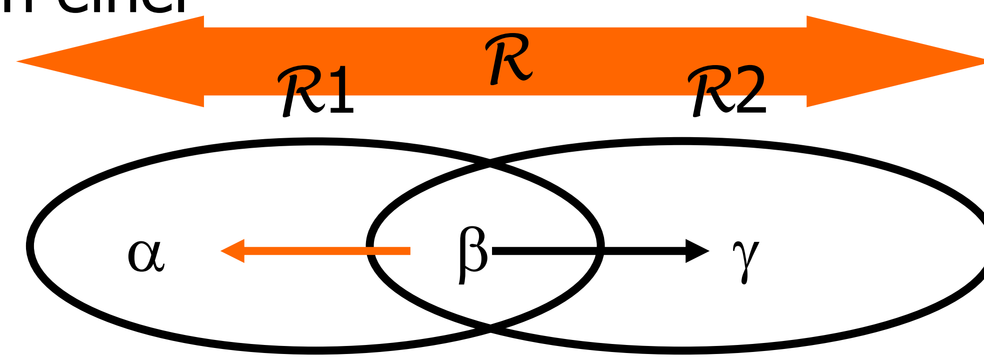
  1. Verlustlosigkeit
     - Die in der ursprünglichen Relationenausprägung R des Schemas $\mathcal{R}$ enthaltenen Informationen müssen aus den Ausprägungen $R1$, ..., Rn der neuen Relationenschemata $\mathcal{R}1$, .., $\mathcal{R}n$ rekonstruierbar sein.

  2. Abhängigkeitserhaltung
     - Die für $\mathcal{R}$ geltenden funktionalen Anhängigkeiten müssen auf die Schemata $\mathcal{R}1$, ..., $\mathcal{R}n$ übertragbar sein.

Kriterien für die Verlustlosigkeit einer Zerlegung

- Die Zerlegung von $\mathcal{R}$ in $\mathcal{R}1$ und $\mathcal{R}2$ ist verlustlos, falls für jede mögliche (gültige) Ausprägung R von $\mathcal{R}$ gilt:
  - $R = R1 \bowtie R2$

- Hinreichende Bedingung für die Verlustlosigkeit einer Zerlegung
  - $(\mathcal{R}1 \cap \mathcal{R}2) \rightarrow \mathcal{R}1$ oder
  - $(\mathcal{R}1 \cap \mathcal{R}2) \rightarrow \mathcal{R}2$
  - Eine Zerlegung von $\mathcal{R}$ in $\mathcal{R}1$ und $\mathcal{R}2$ ist verlustlos, wenn die Joinattribute in einer der Teilrelationen Schlüssel sind.



Biertrinker-Beispiel

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

„Verlustige“ Zerlegung

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

$\Pi_{\text{Kneipe, Gast}}$

$\Pi_{\text{Gast, Bier}}$

Besucht	
Kneipe	*Gast*
Kowalski	Kemper
Kowalski	Eickler
Innsteg	Kemper

Trinkt	
Gast	*Bier*
Kemper	Pils
Eickler	Hefeweizen
Kemper	Hefeweizen

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

$\Pi_{....}$

Besucht	
Kneipe	*Gast*
Kowalski	Kemper
Kowalski	Eickler
Innsteg	Kemper

Trinkt	
Gast	*Bier*
Kemper	Pils
Eickler	Hefeweizen
Kemper	Hefeweizen

$\neq$

$\bowtie$

Besucht A Trinkt		
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 eine „verlustige“ Zerlegung und dementsprechend war die hinreichende Bedingung verletzt. Es gilt nämlich nur die eine nicht-triviale funktionale Abhängigkeit
  - {Kneipe,Gast}→{Bier}

- Wohingegen keine der zwei möglichen, die Verlustlosigkeit garantierenden FDs gelten
  - {Gast}→{Bier}
  - {Gast}→{Kneipe}

- Das liegt daran, dass die Leute (insbes. Kemper) in unterschiedlichen Kneipen unterschiedliches Bier trinken. In derselben Kneipe aber immer das gleiche Bier

Verlustfreie Zerlegung

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

$\Pi_{\text{Vater, Kind}}$

$\Pi_{\text{Mutter, Kind}}$

Väter	
Vater	*Kind*
Johann	Else
Johann	Theo
Heinz	Cleo

Mütter	
Mutter	*Kind*
Martha	Else
Maria	Theo
Martha	Cleo

Erläuterung der verlustfreien Zerlegung der Eltern-Relation

- Eltern: {[Vater, Mutter, Kind]}
- Väter: {[Vater, Kind]}
- Mütter: {[Mutter, Kind]}

- Verlustlosigkeit ist garantiert
- Es gilt nicht nur eine der hinreichenden FDs, sondern gleich beide
  - {Kind}→{Mutter}
  - {Kind}→{Vater}

- Also ist {Kind} natürlich auch der Schlüssel der Relation Eltern

- Die Zerlegung von Eltern ist zwar verlustlos, aber auch ziemlich unnötig, da die Relation in sehr gutem Zustand (~Normalform) ist

Abhängigkeitsbewahrung

- $\mathcal{R}$ ist zerlegt in $\mathcal{R}1, \ldots, \mathcal{R}n$
- $F_{\mathcal{R}} = (F_{\mathcal{R}1} \cup \ldots \cup F_{\mathcal{R}n})$ bzw $F_{\mathcal{R}}+ = (F_{\mathcal{R}1} \cup \ldots \cup F_{\mathcal{R}n})+$

- Beispiel für Abhängigkeitsverlust
  - 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
  - {PLZ} → {Ort, BLand}
  - {Straße, Ort, BLand} → {PLZ}
- Betrachte die Zerlegung
  - 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

$\Pi_{\text{PLZ,Straße}}$

$\Pi_{\text{Stadt,Bland,PLZ}}$

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

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

- Die FD {Straße, Ort, BLand} → {PLZ} ist im zerlegten Schema nicht mehr enthalten → Einfügen inkonsistenter Tupel möglich

Einfügen zweier Tupel, die die FD Ort,Bland,Straße→PLZ verletzen

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

$\Pi_{\text{PLZ,Straße}}$

$\Pi_{\text{Stadt,Bland,PLZ}}$

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

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

Einfügen zweier Tupel, die die FD Ort,Bland,Straße→PLZ verletzen

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

$\bowtie$

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

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

Erste Normalform

- Nur atomare Domänen

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

- 1 NF

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

Zweite Normalform

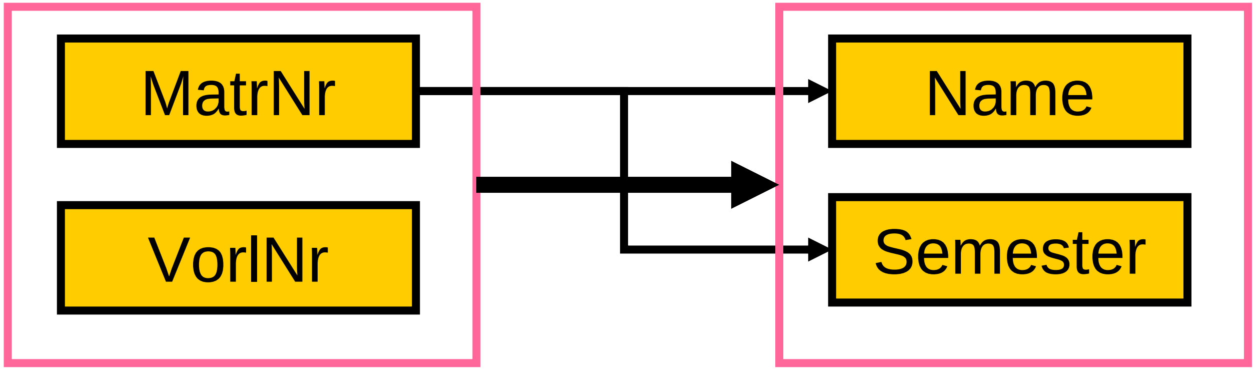
- Eine Relation $\mathcal{R}$ mit zugehörigen FDs $F_{\mathcal{R}}$ ist in zweiter Normalform, falls jedes Nichtschlüssel-Attribut A $\in \mathcal{R}$ voll funktional abhängig ist von jedem Kandidatenschlüssel der Relation.

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
...	...	...	...

Intuitiv: Relationenschema $\mathcal{R}$ verletzt die 2. Normalform, wenn in der Relation Informationen über mehr als ein Konzept modelliert werden.

- Studentenbelegung ist nicht in zweiter NF
  - {MatrNr} → {Name}
  - {MatrNr} → {Semester}

Zweite Normalform



- Einfügeanomalie: Was macht man mit Studenten, die keine Vorlesungenen hören?
- Updateanomalien: Wenn z.B. Carnap ins vierte Semester kommt, muss man sicherstellen, dass alle vier Tupel geändert werden.
- Löschanomalie: Was passiert wenn Fichte ihre einzige Vorlesung absagt?
- Zerlegung in zwei Relationen
  - hören: {[MatrNr, VorlNr]}
  - Studenten: {[MatrNr, Name, Semester]}
- Beide Relationen sind in 2 NF – erfüllen sogar noch „höhere" Gütekriterien ~ Normalformen.

Reicht 2NF?

- Alle Nichtschlüsselattribute müssen vom „ganzen“ Schlüssel abhängig sein

Prüfung						
MatNr	*Name*	*PrüfFachNr*	*PrüfFachName*	*ProfNr*	*ProfName*	*Note*
1	Maier	10	Analysis	20	A	1
1	Maier	11	Algebra	20	A	2
2	Müller	17	DB	40	G	3
2	Müller	18	OOP	40	G	4

2NF

Studenten	
MatNr	*Name*
1	Maier
2	Müller

Prüfung		
MatNr	*PrüfFachNr*	*Note*
1	10	1
1	11	2
2	17	3
2	18	4

PrüfFach			
PrüfFachNr	*PrüfFachName*	*ProfNr*	*ProfName*
10	Analysis	20	A
11	Algebra	30	B
17	DB	40	C
18	OOP	50	D

Reicht 2NF?

- Es gibt noch transitive Abhängigkeiten

Studenten	
MatNr	*Name*
1	Maier
2	Müller

Prüfung		
MatNr	*PrüfFachNr*	*Note*
1	10	1
1	11	2
2	17	3
2	18	4

Prüfung			
PrüfFachNr	*PrüfFachName*	*ProfNr*	*ProfName*
10	Analysis	20	A
11	Algebra	20	A
17	DB	40	G
18	OOP	40	G

- Anomalien z.B.:
- Name eines Professors ändert sich

Dritte Normalform

- Ein Relationenschema $\mathcal{R}$ ist in dritter Normalform, wenn für jede für $\mathcal{R}$ geltende funktionale Abhängigkeit der Form $\alpha \rightarrow B$ mit B $\in$ $\mathcal{R}$ und mindestens eine von drei Bedingungen gilt:
  - B $\in$ α, d.h., die FD ist trivial
  - Das Attribut B ist in einem Kandidatenschlüssel von $\mathcal{R}$ enthalten
  - α ist Superschlüssel von $\mathcal{R}$

3NF Beispiel

Studenten	
MatNr	*Name*
1	Maier
2	Müller

Prüfung		
MatNr	*PrüfFachNr*	*Note*
1	10	1
1	11	2
2	17	3
2	18	4

PrüfFach			
PrüfFachNr	*PrüfFachName*	*ProfNr*	*ProfName*
10	Analysis	20	A
11	Algebra	20	A
17	DB	40	G
18	OOP	40	G

3NF

Studenten	
MatNr	*Name*
1	Maier
2	Müller

Prüfung		
Mat Nr	*PrüfFachNr*	*Note*
1	10	1
1	11	2
2	17	3
2	18	4

PrüfFach		
PrüfFachNr	*PrüfFachName*	*ProfNr*
10	Analysis	20
11	Algebra	20
17	DB	40
18	OOP	40

Prüfer	
ProfNr	*Name*
20	A
40	G

Beispiel 1 3NF Test

- $\mathcal{R}$ = (ABCDEF),
- F = {C → BDAE} = {C → B, C → D, C → A, C → E}.
- Der Schlüssel von $\mathcal{R}$ ist: CF
- ($\mathcal{R}$, F) ist in 3NF, wenn für jede FD eine der drei NF-Bedingungen gilt:

1. Keine FD ist trivial: Bedingung 1 gilt nicht
2. Ist die Attributmenge auf der linken Seite der FDs ein Superschlüssel von $\mathcal{R}$? Nein, Bedingung 2 gilt für keine FD.
3. Ist das Attribut auf der rechten Seite der FDs in einem der Schlüssel von $\mathcal{R}$ enthalten? Nein, Bedingung 3 gilt für keine FD.

- Mindestens eine FD (C → B) verletzt die 3NF. $\mathcal{R}$ ist nicht in 3NF.

Beispiel 2 3NF Test

- $\mathcal{R}$ = (ABCDEF),
- F = {C → B, C → D, D → A, D → E, E → C, E → F, F → E}.
- Schlüssel von $\mathcal{R}$ sind: C, E, F, D
- Keine FD ist trivial: Bedingung 1 gilt nicht
- Ist die Attributmenge auf der linker Seite der FDs ein Superschlüssel von $\mathcal{R}$? OK für
  C → B, C → D, D → A, D → E, E → C, E → F, F → E
- Ist das Attribut auf der rechten Seite der FDs in einem der Schlüssel von R enthalten? OK für
  C → D, D → E, E → C, E → F, F → E gilt aber nicht für
  C → B, D → A
- Keine FD verletzt die 3NF → in 3NF
  (In diesem Fall durch Schritt 2 geklärt)

Zerlegung mit dem Synthesealgorithmus

- Theorem: Ein Relationenschema $\mathcal{R} = \mathcal{R}1 \cup \ldots \cup \mathcal{R}n$ ist in dritter Normalform, wenn alle $\mathcal{R}i$ in dritter Normalform sind.
- Wir geben jetzt einen sogenannten Synthesealgorithmus an, mit dem zu einem gegebenen Relationenschema $\mathcal{R}$ mit funktionalen Anhängigkeiten F eine Zerlegung in $\mathcal{R}1$, ..., $\mathcal{R}n$ ermittelt wird, die alle drei folgenden Kriterien erfüllt.
  - $\mathcal{R}1$, ..., $\mathcal{R}n$ ist eine verlustlose Zerlegung von $\mathcal{R}$.
  - Die Zerlegung $\mathcal{R}1$, ..., $\mathcal{R}n$ ist abhängigkeitserhaltend.
  - Alle $\mathcal{R}1$, ..., $\mathcal{R}n$ sind in dritter Normalform.

Synthesealgorithmus

1. Bestimme die kanonische Überdeckung Fc zu F. Wiederholung:
   a. Linksreduktion
   b. Rechtsreduktion
   c. Entfernung von FDs der Form $\alpha \rightarrow \emptyset$
   d. Zusammenfassung gleicher linker Seiten
2. Für jede funktionale Abhängigkeit $\alpha \rightarrow \beta \in$ Fc:
   - Kreiere ein Relationenschema $\mathcal{R}\alpha := \alpha \cup \beta$
   - Ordne $\mathcal{R}\alpha$ die FDs $F\alpha := \{\alpha` \rightarrow \beta` \in Fc \mid \alpha` \cup \beta` \subseteq \mathcal{R}\alpha\}$ zu.

Informell: aus jeder FD wird ein eigenes Schema (Bedingung für Abhängigkeitstreue erfüllt) - kanonische Überdeckung wichtig, um nicht zu viele oder zu große Teilschemata zu generieren

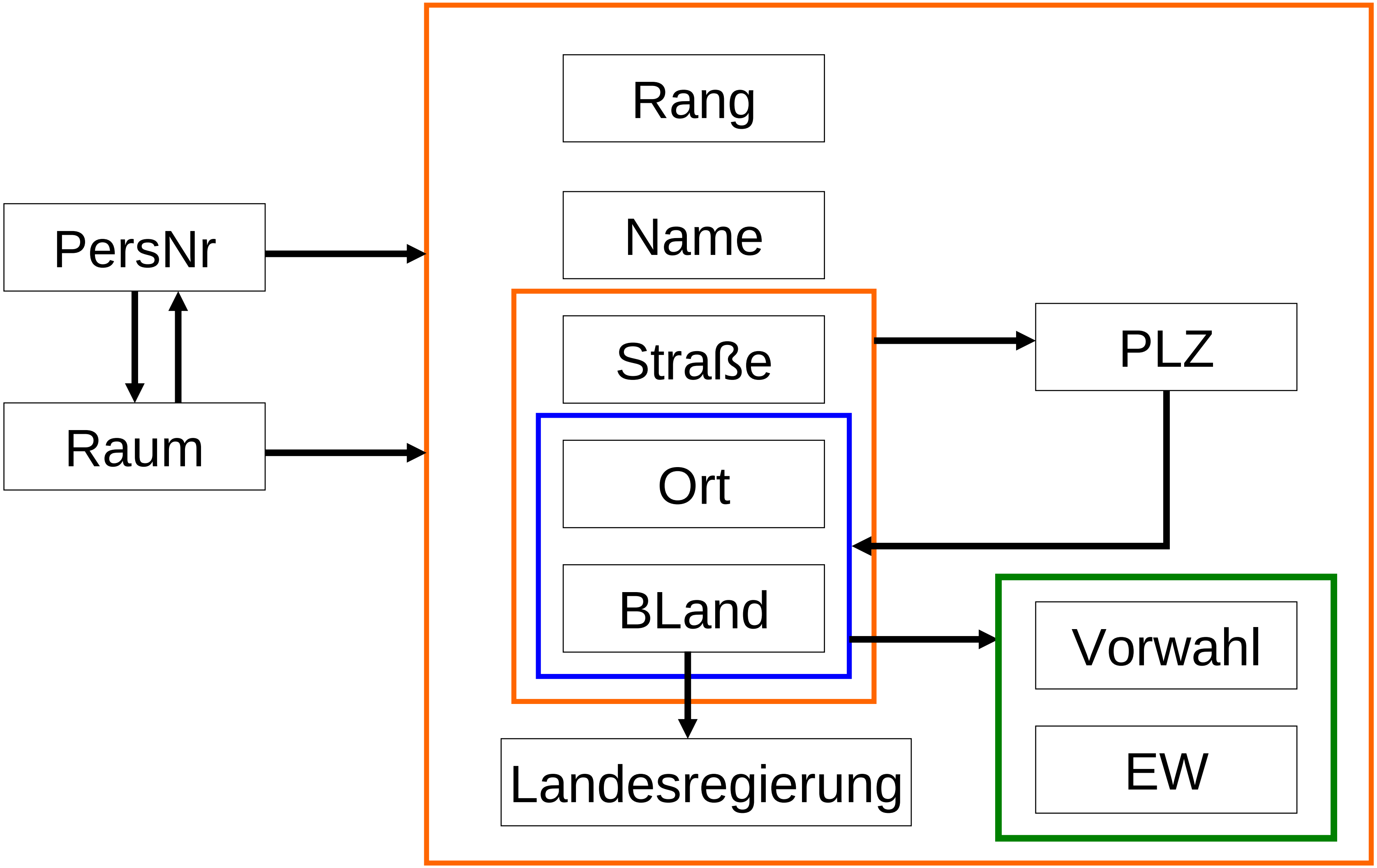
3. Falls eines der in Schritt 2. erzeugten Schemata einen Kandidatenschlüssel von $\mathcal{R}$ bzgl. Fc enthält, sind wir fertig. Sonst wähle einen Kandidatenschlüssel $\kappa \subseteq \mathcal{R}$ aus und definiere folgendes Schema:
   - $\mathcal{R}\kappa := \kappa$
   - $F\kappa := \emptyset$

Informell: erzeuge ein Schema zum Verknüpfen der Teilschemata (Bedingung für Verlustlosigkeit erfüllt)

4. Eliminiere diejenigen Schemata $\mathcal{R}\alpha$, die in einem anderen Relationenschema $\mathcal{R}\alpha`$ enthalten sind, d.h.,
   - $\mathcal{R}\alpha \subseteq \mathcal{R}\alpha`$

Informell: kürze überflüssige Schemata

Anwendung des Synthesealgorithmus



Anwendung des Synthesealgorithmus

ProfessorenAdr: {PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, BLand, EW, Landesregierung} = {P, N, R, Z, O, S, Plz, V, B, E, L}

F = {P → NRZOSB, Z → P, SBO → Plz, OB → EV, B → L, Plz → BO}

1. Kanonische Überdeckung: Hier schon gewährleistet
2. Generierung der Teilschemata und Zuordnung aller FDs:
   - {PNRZOSB} es gelten: P → NRZOSB und Z → P
   - {ZP} es gelten: Z → P und P → Z
   - {SBOPlz} es gelten: SBO → Plz und Plz → BO
   - {OBEV} es gilt: OB → EV
   - {BL} es gilt: B → L
   - {PlzBO} es gilt: Plz → BO

Anwendung des Synthesealgorithmus

3. Enthält eines der Teilschemata einen Schlüssel von $\mathcal{R}$ bezüglich F?
   Ja: P war Schlüssel und ist in {PNRZOSB} enthalten: fertig
4. Schemaelimination:
   - {ZP} ist schon in {PNRZOSB} enthalten: kürzen
   - {PlzBO} ist schon in {SBOPlz} enthalten: kürzen

- Ergebnis:
  - Professoren: {PersNr, Name, Rang, Raum, Ort, Straße, BLand}
  - PLZverzeichnis: {Straße, BLand, Ort, PLZ}
  - OrteVerzeichnis: {Ort, BLand, EW, Vorwahl}
  - Regierungen: {Bland, Landesregierung}

Anwendung des Synthesealgorithmus 2

- $\mathcal{R}$ = (ABCDEF),
- F = {A → EC, BC → F, D → B}.
- Frage: $\mathcal{R}$ in 3NF? Nein. Schlüssel: AD

1. Ist schon kanonisch.
2. $\mathcal{R}_1$ = (AEC), $\mathcal{R}_2$ = (BCF), $\mathcal{R}_3$ = (DB)
   keine (nichttrivialen) Zuordnungen möglich
3. Schlüssel von $\mathcal{R}$: AD → hinzufügen von $\mathcal{R}_4$ = (AD)
4. Nichts zu eliminieren.

$\mathcal{R}$ = (AEC) ∪ (BCF) ∪ (DB) ∪ (AD)

Reicht NF3?

- $\mathcal{R}$ = Städte(Ort, BLand, MP (Ministerpräsident/in), EW (Einwohner))
- F = {Ort BLand → EW, BLand → MP, MP → BLand}.
- Schlüssel : {Ort, BLand}, {Ort, MP} also in 3NF
- „überlappende Schlüsselkandidaten“

Städte			
Ort	***BLand***	***MP***	***EW***
o1	b1	mp1	1
o2	b1	mp1	2
o3	b1	mp1	3
o4	b2	mp2	1

Boyce-Codd-Normalform

- Die Boyce-Codd-Normalform ist eine Weiterentwicklung der Dritten Normalform. In der Dritten Normalform kann es vorkommen, daß ein Teil eines (zusammengesetzten) Schlüsselkandidaten funktional abhängig ist von einem Teil eines anderen Schlüsselkandidaten. Die Boyce-Codd-Normalform verhindert dies.
- Die BCNF braucht nur dann angewendet zu werden, wenn mehrere Schlüsselkandidaten vorhanden sind und sich diese teilweise überlappen. Ist in der Relation nur ein Kandidatenschlüssel vorhanden oder es liegt keine Überlappung bei mehreren Kandidatenschlüsseln vor, befindet sich die Relation automatisch in der BCNF.

Boyce-Codd-Normalform

- Die Boyce-Codd-Normalform (BCNF) ist nochmals eine Verschärfung der 3 NF.
- Ein Relationenschema $\mathcal{R}$ mit FDs F ist in BCNF, wenn für jede für $\mathcal{R}$ geltende funktionale Abhängigkeit der Form $\alpha \to \beta \in F$ und mindestens eine von zwei Bedingungen gilt:
  - $\beta \subseteq \alpha$, d.h., die Abhängigkeit ist trivial oder
  - α ist Superschlüssel von $\mathcal{R}$
- Man kann jede Relation verlustlos in BCNF-Relationen zerlegen
- Manchmal lässt sich dabei die Abhängigkeiterhaltung aber nicht erzielen

Beispiel 1: BCNF Check

- $\mathcal{R}$ = (ABCDEF),
- F = {C → B; C → D;D → A;D → E; E → C; E → F; F → E}.
- Die Schlüssel von $\mathcal{R}$ sind: C, E, D, F
- ($\mathcal{R}$, F) ist in BCNF, wenn für jede FD eine der zwei NF-Bedingungen gilt:

1. Keine FD ist trivial: Bedingung 1 gilt nicht
2. Ist die Attributmenge auf der linken Seite der FDs ein Superschlüssel von $\mathcal{R}$? Ja, Bedingung 2 gilt für alle FDs.

Keine FD verletzt die BCNF → in BCNF

Beispiel 2: BCNF Check

- $\mathcal{R}$ = Städte(Ort, BLand, MP (Ministerpräsident/in), EW (Einwohner))
- F = {Ort BLand → EW, BLand → MP, MP → BLand}.
- Schlüssel : {Ort, BLand}, {Ort, MP}
- ($\mathcal{R}$, F) ist in BCNF, wenn für jede FD eine der zwei NF-Bedingungen gilt:

1. Keine FD ist trivial: Bedingung 1 gilt nicht
2. Ist die Attributmenge auf der linken Seite der FDs ein Superschlüssel von $\mathcal{R}$? Nein, für BLand → MP, MP → BLand.

   $\mathcal{R}$ nicht in BCNF

- Aber: in 3NF, da Bedingung 2 (3NF): auf der rechten Seite steht ein Schlüsselattribut erfüllt ist.

Dekomposition

- Man kann grundsätzlich jedes Relationenschema $\mathcal{R}$ mit funktionalen Anhängigkeiten F so in $\mathcal{R}$1, ..., $\mathcal{R}$n zerlegen, dass gilt:

  - $\mathcal{R}$1, ..., $\mathcal{R}$n ist eine verlustlose Zerlegung von $\mathcal{R}$.

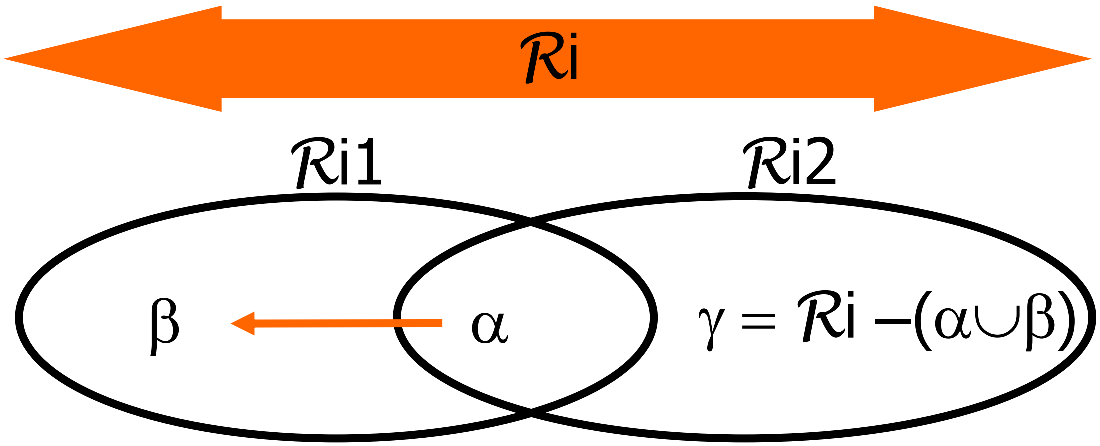
  - Alle $\mathcal{R}$1, ..., $\mathcal{R}$n sind in BCNF.

  - Es kann leider nicht immer erreicht werden, dass die Zerlegung $\mathcal{R}$1, ..., $\mathcal{R}$n abhängigkeitserhaltend ist.

Dekompositions-Algorithmus

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

Veranschaulichung der Dekomposition



Dekomposition der Relation Städte in BCNF-Relationen

- $\mathcal{R}$ = Städte: {[Ort, BLand, MP, EW]}
- F = {BLand → MP, Ort BLand → EW, MP → BLand}
- Schlüssel: {Ort, BLand}, {Ort, MP}
- Z:= $\mathcal{R}$
- BLand → MP verletzt BCNF
  - $\mathcal{R}$i1: Regierungen: {[BLand, MP]} und
    $\mathcal{R}$i2: Städte: {[Ort, BLand, EW]}
  - Z:= {[BLand, MP]} $\cup$ {[Ort, BLand, EW]}

- Zerlegung ist verlustlos und auch abhängigkeitserhaltend

Dekomposition des PLZverzeichnis in BCNF-Relationen

- PLZverzeichnis: {[Straße, Ort, Bland, PLZ]}

- F = {PLZ → Ort Bland, Straße Ort BLand → PLZ}
- Z:= $\mathcal{R}$
- {PLZ → Ort Bland} verletzt BCNF
  - $\mathcal{R}$i1: Regierungen: {PLZ, Ort, BLand} und
    $\mathcal{R}$i2: Städte: {PLZ, Straße}
  - Z:= {PLZ, Ort, BLand} ∪ {PLZ, Straße}

- die Zerlegung ist verlustfrei aber die Abhängigkeit {Straße Ort BLand → PLZ} ist verloren gegangen

Weitere Konzepte

- MVD
  - MultiValued Dependencies / Mehrwertige Abhängigkeiten
  - $\alpha \rightarrow\rightarrow \beta$
  - Einem α ist eine **Menge** von β Werten zugeordnet
- 4NF
  - Ähnlich BCNF nur für MVDs
  - Kann auch mit Dekompositionsalgorithmus hergeleitet werden
- 5NF auch PJNF (Project Join Normalform)
  - Maximale Zerlegung
  - Relation kann nicht mehr durch Join einfacherer Relationen erstellt werden

- Eher theoretische/akademische Bedeutung
- → Wird hier nicht behandelt

Zusammenfassung

- Funktionale Abhängigkeiten
- Transitive Hülle, Kanonische Überdeckung
- Anomalien
- Normalformen
- Synthesealgorithmus
- Dekompositionsalgorithmus