

Relationale Entwurfstheorie

- Funktionale Abhängigkeiten
- Normalformen
- Normalisierung durch Dekomposition



Ziele der relationalen Entwurfstheorie

- Bewertung der Qualität eines Relationenschemas
 - Redundanz
 - Einhaltung von Konsistenzbedingungen
 - Funktionale Abhängigkeiten
- Normalformen als Gütekriterium
- Ggf. Verbesserung eines Relationenschemas
 - durch den Synthesealgorithmus
 - durch Dekomposition

Funktionale Abhängigkeiten

- Def.: Sei \mathcal{R} ein Schema, und $\alpha, \beta \subseteq \mathcal{R}$. Die funktionale Abhängigkeit $\alpha \rightarrow \beta$ gilt, wenn für alle möglichen Ausprägungen R von \mathcal{R} und alle $r, s \in R$ mit $r.\alpha = s.\alpha$ gilt, dass $r.\beta = s.\beta$.
- „Funktionale Abhängigkeit“ wird i.w. mit FD (functional dependency) abgekürzt.
- Bsp.: Schema $\mathcal{R} = \{A, B, C, D\}$ mit einer Ausprägung R :

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

$$\{A\} \rightarrow \{B\}$$

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

$$\text{Nicht: } \{B\} \rightarrow \{C\}$$

Notationskonvention:

$$CD \rightarrow B \text{ statt } \{C, D\} \rightarrow \{B\}$$

Funktionale Abhängigkeiten

- Def.: Sei \mathcal{R} ein Schema, und $\alpha, \beta \subseteq \mathcal{R}$. Die funktionale Abhängigkeit $\alpha \rightarrow \beta$ gilt, wenn für alle möglichen Ausprägungen R von \mathcal{R} und alle $r, s \in R$ mit $r.\alpha = s.\alpha$ gilt, dass $r.\beta = s.\beta$.

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

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
...

- Familie: {[Opa, Oma, Vater, Mutter, Kind]}
- Annahme: [Theo, Martha, Herbert, Maria, Else] bedeutet
 - Theo und Martha sind Eltern von Herbert oder
 - Theo und Martha sind Eltern von Maria

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

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
 - $\alpha \rightarrow \beta$ gilt und
 - α nicht mehr verkleinert werden kann,
 - d.h. für kein $A \in \alpha$ gilt $(\alpha - \{A\}) \rightarrow \beta$.
 - Notation für volle funktionale Abhängigkeit: $\alpha \rightarrow \cdot \beta$
- $\alpha \subseteq \mathcal{R}$ ist ein Kandidaten-Schlüssel, falls $\alpha \rightarrow \cdot \mathcal{R}$ gilt.

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*:

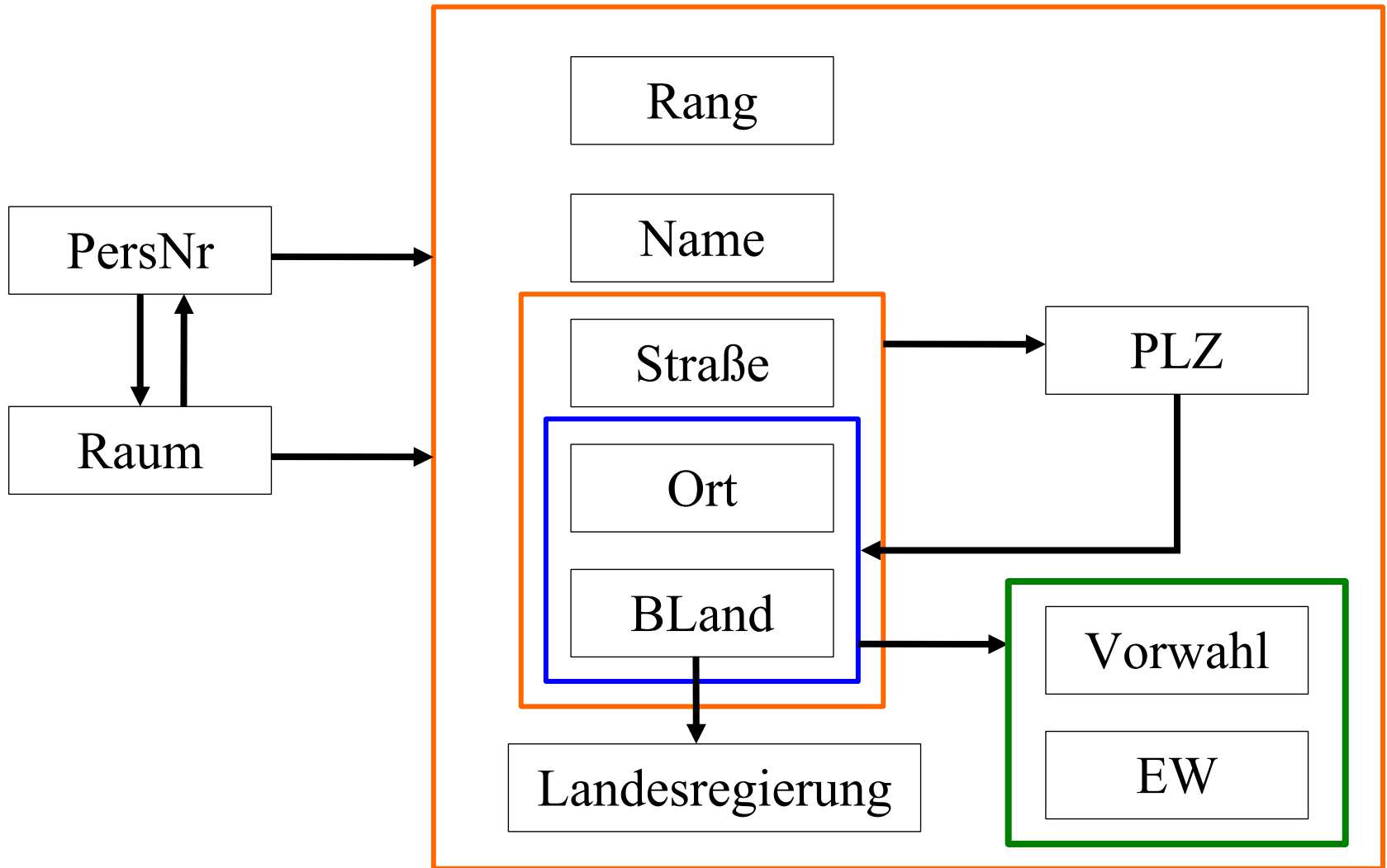
- {Name, BLand}
- {Name, Vorwahl}

- Beachte, dass zwei kleinere Städte dieselbe Vorwahl haben können

Bestimmung funktionaler Abhängigkeiten

- 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



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 $\alpha\gamma \rightarrow \beta\gamma$. Hierbei stehe z.B. $\alpha\gamma$ für $\alpha \cup \gamma$.
- Transitivität
 - Falls $\alpha \rightarrow \beta$ und $\beta \rightarrow \gamma$ gilt, dann gilt auch $\alpha \rightarrow \gamma$.
- 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 $\alpha\gamma \rightarrow \delta$

Bestimmung der Hülle einer Attributmeng

- Eingabe: eine Menge F von FDs, eine Menge α von Attributen.
- Ausgabe: die größte Menge α^+ von Attributen, für die $\alpha \rightarrow \alpha^+$ gilt.

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

Kanonische Überdeckung

- F_c heißt kanonische Überdeckung von F , wenn die folgenden drei Kriterien erfüllt sind:
 1. $F_c \equiv F$, d.h. $F_c^+ = F^+$
 2. In F_c existieren keine FDs, die überflüssige Attribute enthalten. D.h. es muß folgendes gelten:
 - $\forall A \in \alpha: (F_c - (\alpha \rightarrow \beta) \cup ((\alpha - \{A\}) \rightarrow \beta)) \neq F_c$
 - $\forall B \in \beta: (F_c - (\alpha \rightarrow \beta) \cup (\alpha \rightarrow (\beta - \{B\}))) \neq F_c$
 3. Jede linke Seite einer funktionalen Abhängigkeit in F_c 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:
 - Überprüfe für alle $A \in \alpha$, ob A überflüssig ist, d.h., ob
 - $\beta \subseteq \text{AttrHülle}(F, \alpha - A)$
 - gilt. Falls dies der Fall ist, ersetze $\alpha \rightarrow \beta$ durch $(\alpha - A) \rightarrow \beta$.
2. Führe für jede (verbliebene) FD die Rechtsreduktion durch:
 - Überprüfe für alle $B \in \beta$, ob
 - $B \in \text{AttrHülle}(F - (\alpha \rightarrow \beta) \cup (\alpha \rightarrow (\beta - B)), \alpha)$
 - 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 FDs der Form $\alpha \rightarrow \beta_1, \dots, \alpha \rightarrow \beta_n$ zusammen, so dass $\alpha \rightarrow (\beta_1 \cup \dots \cup \beta_n)$ verbleibt.

Kanonische Überdeckung

Frage: Warum ist die kanonische Überdeckung interessant?

- Antwort a): Sie wird weiter unten im Synthesealgorithmus zur Berechnung der 3NF benötigt.
- Antwort b): In einer DB müssen bei jedem Update alle Constraints überprüft werden. Je kompakter wir die FDs beschreiben können (ohne Information zu verlieren), desto effizienter ist dies.

Bemerkung: Es kann verschiedene kanonische Überdeckungen zu einer Menge von FDs geben.

Übung: Berechnen Sie alle kanonischen Überdeckungen von $F = \{A \rightarrow B, B \rightarrow A, AB \rightarrow CD, C \rightarrow D, D \rightarrow C\}$

„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

- Es gibt zwei Korrektheitskriterien für die Zerlegung von Relationenschemata:

1. Verlustlosigkeit

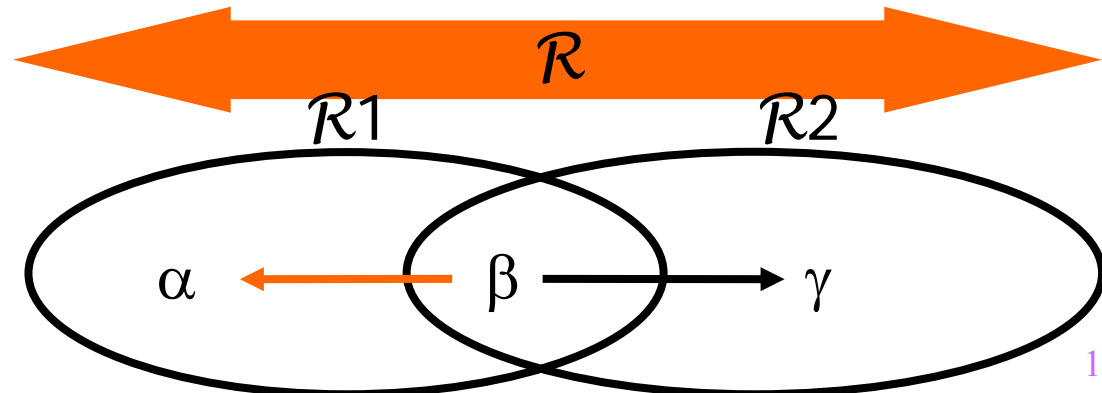
- Die in der ursprünglichen Relationenausprägung R des Schemas \mathcal{R} enthaltenen Informationen müssen aus den Ausprägungen R_1, \dots, R_n der neuen Relationenschemata $\mathcal{R}_1, \dots, \mathcal{R}_n$ rekonstruierbar sein.

1. Abhängigkeitserhaltung

- Die für \mathcal{R} geltenden funktionalen Abhängigkeiten müssen auf die Schemata $\mathcal{R}_1, \dots, \mathcal{R}_n$ übertragbar sein.

Kriterien für die Verlustlosigkeit einer Zerlegung

- $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$
 - $\mathcal{R}_1 := \Pi_{\mathcal{R}_1}(\mathcal{R})$
 - $\mathcal{R}_2 := \Pi_{\mathcal{R}_2}(\mathcal{R})$
- 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:
 - $R = R_1 \bowtie R_2$
- Hinreichende Bedingung für die Verlustlosigkeit einer Zerlegung
 - $(\mathcal{R}_1 \cap \mathcal{R}_2) \rightarrow \mathcal{R}_1$ oder
 - $(\mathcal{R}_1 \cap \mathcal{R}_2) \rightarrow \mathcal{R}_2$



Biertrinker-Beispiel

<i>Biertrinker</i>		
<i>Kneipe</i>	<i>Gast</i>	<i>Bier</i>
Kowalski	Kemper	Pils
Kowalski	Eickler	Hefeweizen
Innsteg	Kemper	Hefeweizen

„Verlustige“ Zerlegung

<i>Biertrinker</i>		
<i>Kneipe</i>	<i>Gast</i>	<i>Bier</i>
Kowalski	Kemper	Pils
Kowalski	Eickler	Hefeweizen
Innsteg	Kemper	Hefeweizen

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

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

<i>Besucht</i>	
<i>Kneipe</i>	<i>Gast</i>
Kowalski	Kemper
Kowalski	Eickler
Innsteg	Kemper

<i>Trinkt</i>	
<i>Gast</i>	<i>Bier</i>
Kemper	Pils
Eickler	Hefeweizen
Kemper	Hefeweizen

<i>Biertrinker</i>		
<i>Kneipe</i>	<i>Gast</i>	<i>Bier</i>
Kowalski	Kemper	Pils
Kowalski	Eickler	Hefeweizen
Innsteg	Kemper	Hefeweizen

<i>Besucht</i>	
<i>Kneipe</i>	<i>Gast</i>
Kowalski	Kemper
Kowalski	Eickler
Innsteg	Kemper

<i>Trinkt</i>	
<i>Gast</i>	<i>Bier</i>
Kemper	Pils
Eickler	Hefeweizen
Kemper	Hefeweizen

$\Pi \dots$

\bowtie

<i>Besucht \bowtie Trinkt</i>		
<i>Kneipe</i>	<i>Gast</i>	<i>Bier</i>
Kowalski	Kemper	Pils
Kowalski	Kemper	Hefeweizen
Kowalski	Eickler	Hefeweizen
Innsteg	Kemper	Pils
Innsteg	Kemper	Hefeweizen

\neq

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
 - $\{\text{Kneipe}, \text{Gast}\} \rightarrow \{\text{Bier}\}$
- Wohingegen keine der zwei möglichen, die Verlustlosigkeit garantierenden FDs gelten
 - $\{\text{Gast}\} \rightarrow \{\text{Bier}\}$
 - $\{\text{Gast}\} \rightarrow \{\text{Kneipe}\}$
- Das liegt daran, dass die Leute (insbes. Kemper) in unterschiedlichen Kneipen unterschiedliches Bier trinken. In derselben Kneipe aber immer das gleiche Bier
 - (damit sich die KellnerInnen darauf einstellen können?)

Verlustfreie Zerlegung

<i>Eltern</i>		
<i>Vater</i>	<i>Mutter</i>	<i>Kind</i>
Johann	Martha	Else
Johann	Maria	Theo
Heinz	Martha	Cleo

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

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

<i>Väter</i>	
<i>Vater</i>	<i>Kind</i>
Johann	Else
Johann	Theo
Heinz	Cleo

<i>Mütter</i>	
<i>Mutter</i>	<i>Kind</i>
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 $\mathcal{R}_1, \dots, \mathcal{R}_n$
- $F_{\mathcal{R}} = (F_{\mathcal{R}_1} \cup \dots \cup F_{\mathcal{R}_n})$ bzw. $F_{\mathcal{R}^+} = (F_{\mathcal{R}_1} \cup \dots \cup F_{\mathcal{R}_n})^+$
- Beispiel für Abhängigkeitsverlust
 - PLZverzeichnis: $\{[\text{Straße}, \text{Ort}, \text{Bland}, \text{PLZ}]\}$
- 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
 - $\{\text{PLZ}\} \rightarrow \{\text{Ort}, \text{BLand}\}$
 - $\{\text{Straße}, \text{Ort}, \text{BLand}\} \rightarrow \{\text{PLZ}\}$
- Betrachte die Zerlegung
 - Straßen: $\{[\text{PLZ}, \text{Straße}]\}$
 - Orte: $\{[\text{PLZ}, \text{Ort}, \text{BLand}]\}$

Zerlegung der Relation PLZverzeichnis

<i>PLZverzeichnis</i>			
<i>Ort</i>	<i>BLand</i>	<i>Straße</i>	<i>PLZ</i>
Frankfurt	Hessen	Goethestraße	60313
Frankfurt	Hessen	Galgenstraße	60437
Frankfurt	Brandenburg	Goethestraße	15234

$\Pi_{PLZ, Straße}$

<i>Straßen</i>	
<i>PLZ</i>	<i>Straße</i>
15234	Goethestraße
60313	Goethestraße
60437	Galgenstraße

$\Pi_{Stadt, Bland, PLZ}$

<i>Orte</i>		
<i>Ort</i>	<i>BLand</i>	<i>PLZ</i>
Frankfurt	Hessen	60313
Frankfurt	Hessen	60437
Frankfurt	Brandenburg	15234

• Die FD $\{Straße, Ort, BLand\} \rightarrow \{PLZ\}$ ist im zerlegten Schema nicht mehr enthalten \rightarrow Einfügen inkonsistenter Tupel möglich

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

<i>PLZverzeichnis</i>			
<i>Ort</i>	<i>BLand</i>	<i>Straße</i>	<i>PLZ</i>
Frankfurt	Hessen	Goethestraße	60313
Frankfurt	Hessen	Galgenstraße	60437
Frankfurt	Brandenburg	Goethestraße	15234

$\Pi_{PLZ, Straße}$

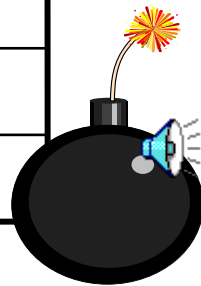
$\Pi_{Stadt, Bland, PLZ}$

<i>Straßen</i>	
<i>PLZ</i>	<i>Straße</i>
15234	Goethestraße
60313	Goethestraße
60437	Galgenstraße
15235	Goethestrasse

<i>Orte</i>		
<i>Ort</i>	<i>BLand</i>	<i>PLZ</i>
Frankfurt	Hessen	60313
Frankfurt	Hessen	60437
Frankfurt	Brandenburg	15234
Frankfurt	Brandenburg	15235

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

<i>PLZverzeichnis</i>			
<i>Ort</i>	<i>BLand</i>	<i>Straße</i>	<i>PLZ</i>
Frankfurt	Hessen	Goethestraße	60313
Frankfurt	Hessen	Galgenstraße	60437
Frankfurt	Brandenburg	Goethestraße	15234
Frankfurt	Brandenburg	Goethestraße	15235



<i>Straßen</i>	
<i>PLZ</i>	<i>Straße</i>
15234	Goethestraße
60313	Goethestraße
60437	Galgenstraße
15235	Goethestrasse

<i>Orte</i>		
<i>Ort</i>	<i>BLand</i>	<i>PLZ</i>
Frankfurt	Hessen	60313
Frankfurt	Hessen	60437
Frankfurt	Brandenburg	15234
Frankfurt	Brandenburg	15235

Erste Normalform

- Nur atomare Domänen

<i>Eltern</i>		
<i>Vater</i>	<i>Mutter</i>	<i>Kinder</i>
Johann	Martha	{Else, Lucie}
Johann	Maria	{Theo, Josef}
Heinz	Martha	{Cleo}

- 1 NF

<i>Eltern</i>		
<i>Vater</i>	<i>Mutter</i>	<i>Kind</i>
Johann	Martha	Else
Johann	Martha	Lucie
Johann	Maria	Theo
Johann	Maria	Josef
Heinz	Martha	Cleo

Exkurs: NF²-Relationen

- Non-First Normal-Form-Relationen
- Geschachtelte Relationen

<i>Eltern</i>			
<i>Vater</i>	<i>Mutter</i>	<i>Kinder</i>	
		<i>KName</i>	<i>KAlter</i>
Johann	Martha	Else	5
		Lucie	3
Johann	Maria	Theo	3
		Josef	1
Heinz	Martha	Cleo	9

Zweite Normalform

- In einer Relation soll immer nur ein Konzept modelliert werden.
- Jedes Nichtschlüssel-Attribut soll einen Fakt zu dem durch den gesamten(!) Schlüssel identifizierten Konzept beschreiben.
- Def.: 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.

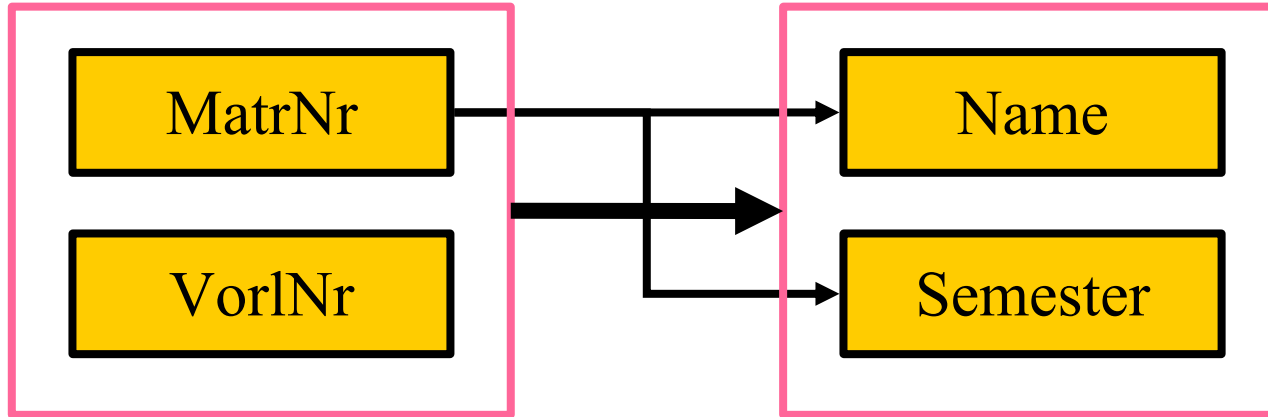
Zweite Normalform

- Def. (Wdh.): 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
...

- Studentenbelegung ist nicht in zweiter NF
 - $\{\text{MatrNr}\} \rightarrow \{\text{Name}\}$
 - $\{\text{MatrNr}\} \rightarrow \{\text{Semester}\}$

Zweite Normalform



- 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 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.

Dritte Normalform

- Derselbe Fakt soll nicht mehrfach gespeichert werden.
- D.h., kein Nichtschlüssel-Attribut soll einen Fakt einer Attributmenge darstellen, die kein Schlüssel ist.
- Def.: Ein Relationenschema \mathcal{R} ist in dritter Normalform, wenn für jede für \mathcal{R} geltende funktionale Abhängigkeit der Form $\alpha \rightarrow B$ mit $B \in \mathcal{R}$ mindestens **eine** von drei Bedingungen gilt:
 - $B \in \alpha$, d.h., die FD ist trivial.
 - α ist Superschlüssel von \mathcal{R} .
 - Das Attribut B ist in einem Kandidatenschlüssel von \mathcal{R} enthalten – d.h. B ist *prim*.

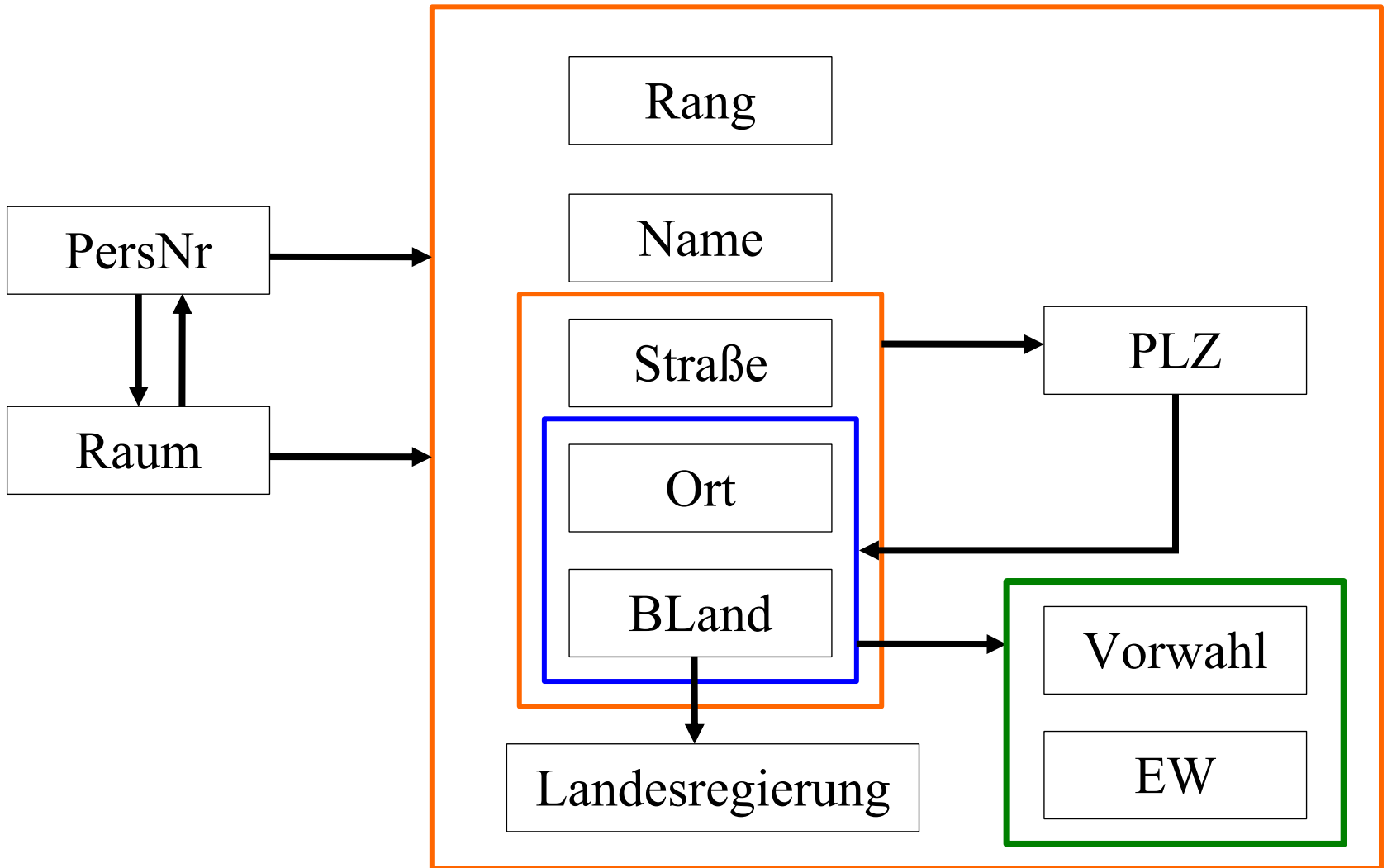
Beispiel

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

mit den 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}

ist nicht in 3NF, da z.B. {Ort, BLand} \rightarrow {Vorwahl} keine der drei Bedingungen der Definition erfüllt.



Zerlegung mit dem Synthesealgorithmus

- Der Synthesealgorithmus ermittelt zu einem gegebenen Relationenschema \mathcal{R} mit funktionalen Abhängigkeiten F eine Zerlegung in $\mathcal{R}_1, \dots, \mathcal{R}_n$, die die drei folgenden Kriterien erfüllt:
 - $\mathcal{R}_1, \dots, \mathcal{R}_n$ ist eine verlustlose Zerlegung von \mathcal{R} .
 - Die Zerlegung $\mathcal{R}_1, \dots, \mathcal{R}_n$ ist abhängigkeiterhaltend.
 - Alle $\mathcal{R}_1, \dots, \mathcal{R}_n$ sind in dritter Normalform.

Synthesealgorithmus

1. Bestimme die kanonische Überdeckung F_c 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 $\alpha \rightarrow \beta \in F_c$:

- Kreiere ein Relationenschema $\mathcal{R}_\alpha := \alpha \cup \beta$
- Ordne \mathcal{R}_α die FDs $F_\alpha := \{\alpha' \rightarrow \beta' \in F_c \mid \alpha' \cup \beta' \subseteq \mathcal{R}_\alpha\}$ zu.

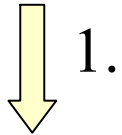
3. Falls eines der in Schritt 2. erzeugten Schemata einen Kandidatenschlüssel von \mathcal{R} bzgl. F_c enthält, sind wir fertig. Sonst wähle einen Kandidatenschlüssel $\kappa \subseteq \mathcal{R}$ aus und definiere zusätzlich folgendes Schema:

- $\mathcal{R}_\kappa := \kappa$
- $F_\kappa := \emptyset$

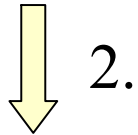
4. Eliminiere diejenigen Schemata \mathcal{R}_α , die in einem anderen Relationenschema $\mathcal{R}_{\alpha'}$ enthalten sind, d.h. für die $\mathcal{R}_\alpha \subseteq \mathcal{R}_{\alpha'}$ gilt.

Anwendung des Synthesealgorithmus

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

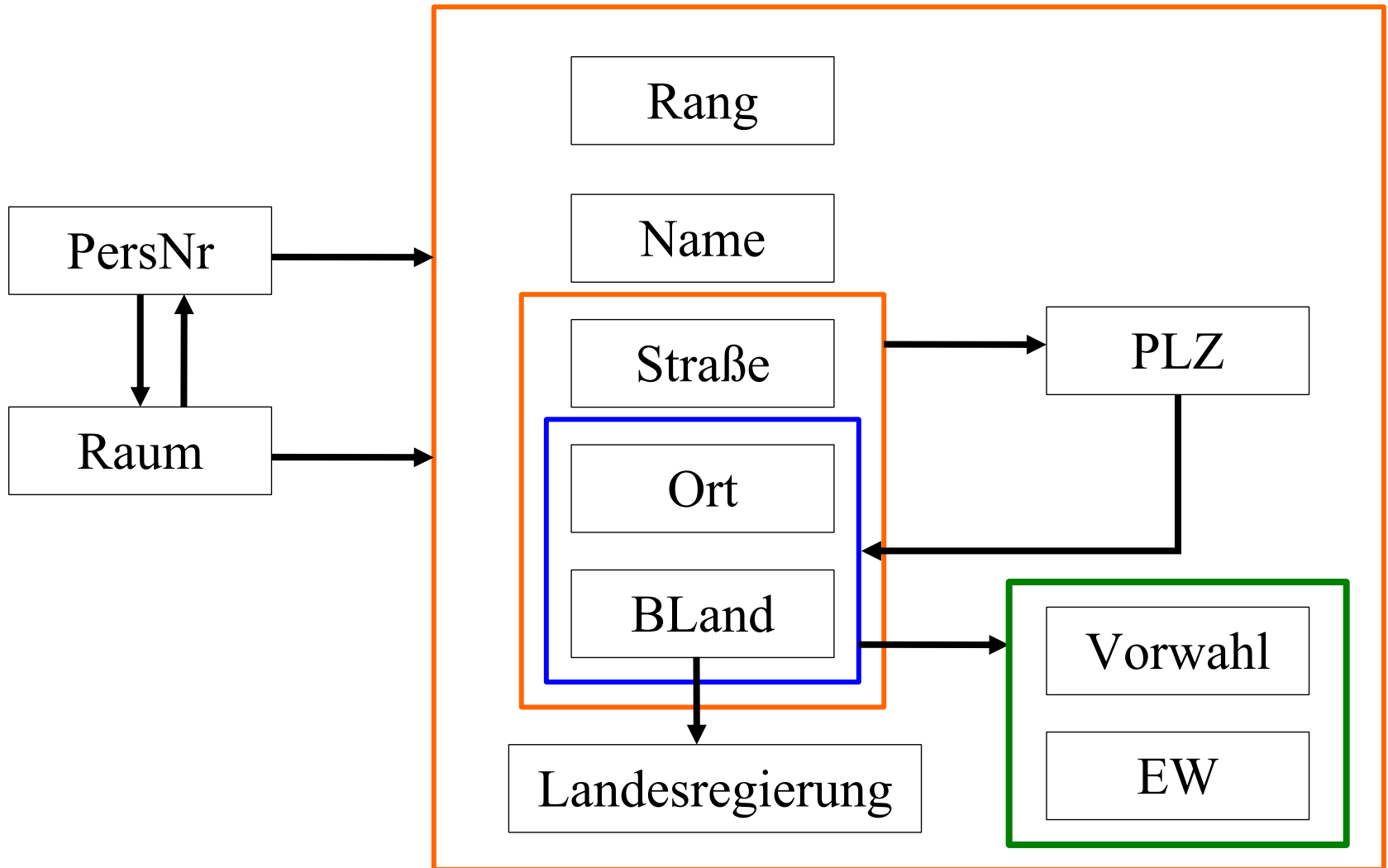


1. {PersNr} → {Name, Rang, Raum, Ort, Straße, BLand}
2. {Raum} → {PersNr}
3. {Straße, BLand, Ort} → {PLZ}
4. {Ort, BLand} → {EW, Vorwahl}
5. {BLand} → {Landesregierung}
6. {PLZ} → {BLand, Ort}



- Professoren: {[PersNr, Name, Rang, Raum, Ort, Straße, BLand]}
- PLZverzeichnis: {[Straße, BLand, Ort, PLZ]}
- Orteverzeichnis: {[Ort, BLand, EW, Vorwahl]}
- Regierungen: {[BLand, Landesregierung]}

Anwendung des Synthesealgorithmus



Boyce-Codd-Normalform

- Die Boyce-Codd-Normalform (BCNF) ist eine Verschärfung der 3 NF. Sie soll vermeiden, dass Information mehrfach abgespeichert wird.
- Ein Relationenschema \mathcal{R} mit FDs F ist in BCNF, wenn für jede für \mathcal{R} geltende funktionale Abhängigkeit der Form $\alpha \rightarrow \beta \in F$ mindestens **eine** der zwei Bedingungen gilt:
 - $\beta \subseteq \alpha$, d.h., die Abhängigkeit ist trivial
 - α ist Superschlüssel von \mathcal{R} .
- Man kann jede Relation **verlustlos** in BCNF-Relationen zerlegen.
- Manchmal lässt sich dabei die **Abhängigkeitserhaltung** aber **nicht** erzielen. (Deswegen ist neben BCNF auch die 3NF noch interessant.)

Städte ist in 3NF, aber nicht in BCNF

- Städte: {[Ort, BLand, Ministerpräsident/in, EW]}
- Geltende FDs:
 - {Ort, BLand} → {EW}
 - {BLand} → {Ministerpräsident/in}
 - {Ministerpräsident/in} → {BLand}
- Schlüsselkandidaten:
 - {Ort, BLand}
 - {Ort, Ministerpräsident/in}

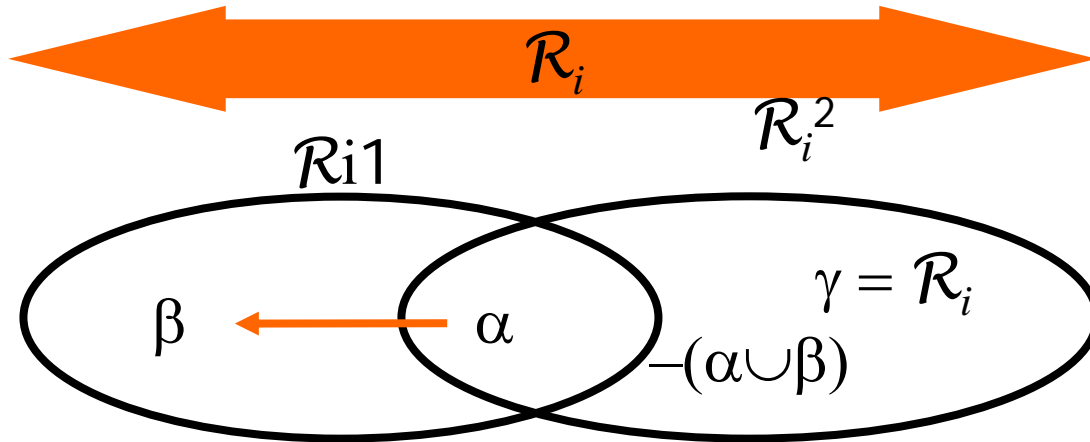
Dekomposition

- Man kann grundsätzlich jedes Relationenschema \mathcal{R} mit funktionalen Anhängigkeiten F so in $\mathcal{R}_1, \dots, \mathcal{R}_n$ zerlegen, dass gilt:
 - $\mathcal{R}_1, \dots, \mathcal{R}_n$ ist eine verlustlose Zerlegung von \mathcal{R} .
 - Alle $\mathcal{R}_1, \dots, \mathcal{R}_n$ sind in BCNF.
 - Es kann leider nicht immer erreicht werden, dass die Zerlegung $\mathcal{R}_1, \dots, \mathcal{R}_n$ abhängigkeiterhaltend 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
 - $\alpha \cap \beta = \emptyset$
 - $\neg(\alpha \rightarrow \mathcal{R}_i)$
 - Finde eine solche FD.
 - Man sollte sie so wählen, dass β alle von α funktional abhängigen Attribute $B \in (\mathcal{R}_i - \alpha)$ enthält, damit der Dekompositionsalgorithmus möglichst schnell terminiert.
 - Zerlege \mathcal{R}_i in $\mathcal{R}_i^1 := \alpha \cup \beta$ und $\mathcal{R}_i^2 := \mathcal{R}_i - \beta$
 - Entferne \mathcal{R}_i aus Z und füge \mathcal{R}_i^1 und \mathcal{R}_i^2 ein, also
 - $Z := (Z - \{\mathcal{R}_i\}) \cup \{\mathcal{R}_i^1\} \cup \{\mathcal{R}_i^2\}$

Veranschaulichung der Dekomposition

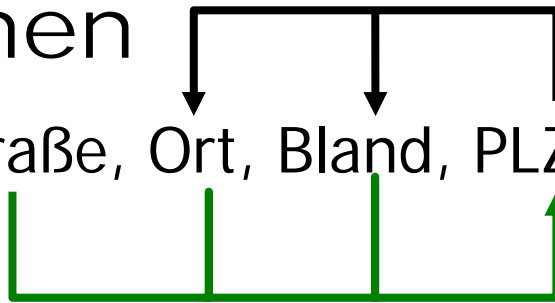


Dekomposition der Relation Städte in BCNF-Relationen

- Städte: {[Ort, BLand, Ministerpräsident/in, EW]}
- Geltende FDs:
 - {BLand} \rightarrow {Ministerpräsident/in}
 - {Ort, BLand} \rightarrow {EW}
 - {Ministerpräsident/in} \rightarrow {BLand}
- \mathcal{R}_1^1 :
 - Regierungen: {[BLand, Ministerpräsident/in]}
- \mathcal{R}_1^2 :
 - Städte: {[Ort, BLand, EW]}
- Diese Zerlegung ist verlustlos (und auch abhängigkeiterhaltend)

Dekomposition des PLZverzeichnis in BCNF-Relationen

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



- Funktionale Abhängigkeiten:
 - {PLZ} → {Ort, Bland}
 - {Straße, Ort, Bland} → {PLZ}
- Betrachte die Zerlegung
 - Straßen: {[PLZ, Straße]}
 - Orte: {[PLZ, Ort, Bland]}
- Diese Zerlegung
 - ist verlustlos,
 - aber nicht abhängigkeiterhaltend (siehe oben).

Mehrwertige Abhängigkeiten (MVDs)

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

- Def.: Die mehrwertige Abhängigkeit $\alpha \twoheadrightarrow \beta$ gilt, wenn aus der Existenz zweier Tupel t_1 und t_2 mit gleichen α -Werten die Existenz zweier Tupel t_3 und t_4 folgt mit
 - $t_3.\alpha = t_4.\alpha = t_1.\alpha = t_2.\alpha$
 - $t_3.\beta = t_1.\beta, t_4.\beta = t_2.\beta$
 - $t_3.\gamma = t_2.\gamma, t_4.\gamma = t_1.\gamma$

Many-Valued Dependencies (MVDs)

- Jede FD ist auch eine MVD.
- MVDs heissen auch Tuple-generating dependencies:
 - Man kann eine Relation MVD-konform machen, indem man zusätzliche Tupel einfügt. (Bei FDs geht das nicht.)

Mehrwertige Abhängigkeiten

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

● $A \twoheadrightarrow B$

● $A \twoheadrightarrow C$

Mehrwertige Abhängigkeiten: ein 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 Abhängigkeiten: ein Beispiel

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

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

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

Sprachen	
PersNr	Sprache
3002	griechisch
3002	lateinisch
30005	deutsch

Sprachen	
PersNr	ProgSprache
3002	C
3002	Pascal
30005	Ada

Mehrwertige Abhängigkeiten: ein Beispiel

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

Sprachen	
PersNr	Sprache
3002	griechisch
3002	lateinisch
30005	deutsch

Sprachen	
PersNr	ProgSprache
3002	C
3002	Pascal
30005	Ada



Beispiel (Fortsetzung)

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
...

- Familie: {[Opa, Oma, Vater, Mutter, Kind]}
- Annahme: [Theo, Martha, Herbert, Maria, Else] bedeutet
 - Theo und Martha sind Eltern von Herbert oder
 - Theo und Martha sind Eltern von Maria

Beispiel (Fortsetzung)

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 \rightarrow Vater, Mutter
- Kind, Opa \rightarrow Oma
- Kind, Oma \rightarrow Opa

MVDs ?

Beispiel (Fortsetzung)

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 \rightarrow Vater, Mutter
- Kind, Opa \rightarrow Oma
- Kind, Oma \rightarrow Opa
- V, M $\rightarrow\rightarrow$ K
- V, M $\rightarrow\rightarrow$ Opa, Oma

Verlustlose Zerlegung bei MVDs: hinreichende + notwendige Bedingung

- $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$
 - $R_1 := \Pi_{\mathcal{R}_1} (R)$
 - $R_2 := \Pi_{\mathcal{R}_2} (R)$
- 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:
 - $R = R_1 \bowtie R_2$
- Die Zerlegung von \mathcal{R} in \mathcal{R}_1 und \mathcal{R}_2 ist verlustlos **genau dann wenn**
 - $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$
 - und** mindestens eine von zwei MVDs gilt:
 - $(\mathcal{R}_1 \cap \mathcal{R}_2) \twoheadrightarrow \mathcal{R}_1$ oder
 - $(\mathcal{R}_1 \cap \mathcal{R}_2) \twoheadrightarrow \mathcal{R}_2$

Inferenzregeln für MVDs

- *Reflexivität*: $\beta \subseteq \alpha \Rightarrow \alpha \rightarrow \beta$
- *Verstärkung*: Sei $\alpha \rightarrow \beta$. Dann gilt $\gamma\alpha \rightarrow \gamma\beta$.
- *Transitivität*: Sei $\alpha \rightarrow \beta$ und $\beta \rightarrow \gamma$. Dann gilt $\alpha \rightarrow \gamma$.
- *Komplement*: $\alpha \twoheadrightarrow \beta$. Dann gilt $\alpha \twoheadrightarrow \mathcal{R} - \beta - \alpha$.
- *Mehrwertige Verstärkung*: Sei $\alpha \twoheadrightarrow \beta$ und $\delta \subseteq \gamma$. Dann gilt $\gamma\alpha \twoheadrightarrow \delta\beta$.
- *Mehrwertige Transitivität*: Sei $\alpha \twoheadrightarrow \beta$ und $\beta \twoheadrightarrow \gamma$. Dann gilt $\alpha \twoheadrightarrow \gamma - \beta$.
- *Verallgemeinerung*: Sei $\alpha \rightarrow \beta$. Dann gilt $\alpha \twoheadrightarrow \beta$.

Inferenzregeln für MVDs (Forts.)

- *Koaleszenz*: Sei $\alpha \twoheadrightarrow \beta$ und $\gamma \subseteq \beta$. Existiert ein $\delta \subseteq \mathcal{R}$, so daß $\delta \cap \beta = \emptyset$ und $\delta \rightarrow \gamma$, gilt $\alpha \rightarrow \gamma$.
- *Mehrwertige Vereinigung*: sei $\alpha \twoheadrightarrow \beta$ und $\alpha \twoheadrightarrow \gamma$. Dann gilt $\alpha \twoheadrightarrow \gamma\beta$.
- *Schnittmenge*: Sei $\alpha \twoheadrightarrow \beta$ und $\alpha \twoheadrightarrow \gamma$. Dann gilt $\alpha \twoheadrightarrow \beta \cap \gamma$.
- *Differenz*: Sei $\alpha \twoheadrightarrow \beta$ und $\alpha \twoheadrightarrow \gamma$. Dann gilt $\alpha \twoheadrightarrow \beta - \gamma$ und $\alpha \twoheadrightarrow \gamma - \beta$.

Triviale MVDs

- Def.: Eine MVD ist trivial, wenn sie von jeder Relationenausprägung erfüllt wird.
- Satz: Eine MVD $\alpha \twoheadrightarrow \beta$ ist trivial genau dann wenn
 - $\beta \subseteq \alpha$ oder
 - $\beta = R - \alpha$.

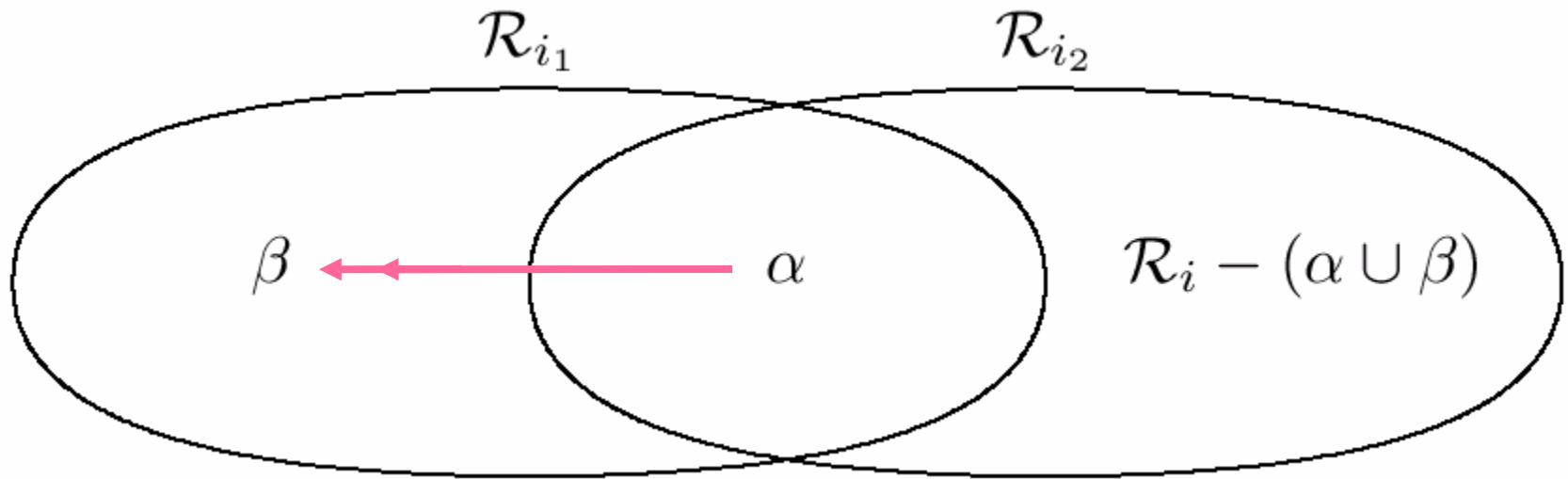
4. Normalform

- Def.: Eine Relation \mathcal{R} ist in 4 NF, wenn für jede MVD $\alpha \twoheadrightarrow \beta$ eine der folgenden Bedingungen gilt:
 - Die MVD ist trivial **oder**
 - α ist Superschlüssel von \mathcal{R} .

Dekomposition in 4 NF

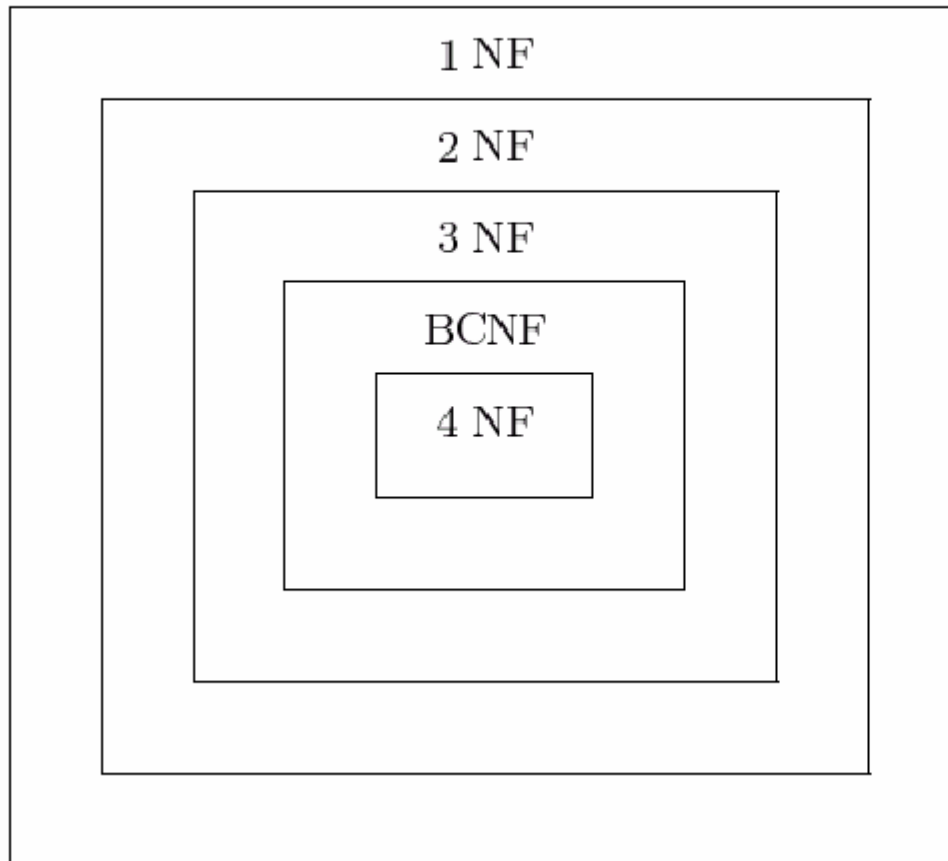
- Starte mit der Menge $Z := \{\mathcal{R}\}$
- Solange es noch ein Relationenschema \mathcal{R}_i in Z gibt, das nicht in 4NF ist, mache folgendes:
 - Es gibt also eine für \mathcal{R}_i geltende nicht-triviale MVD $(\alpha \twoheadrightarrow \beta)$, für die gilt:
 - $\alpha \cap \beta = \emptyset$
 - $\neg(\alpha \rightarrow \mathcal{R}_i)$
 - Finde eine solche MVD
 - Zerlege \mathcal{R}_i in $\mathcal{R}_i^1 := \alpha \cup \beta$ und $\mathcal{R}_i^2 := \mathcal{R}_i - \beta$
 - Entferne \mathcal{R}_i aus Z und füge \mathcal{R}_i^1 und \mathcal{R}_i^2 ein, also
 - $Z := (Z - \{\mathcal{R}_i\}) \cup \{\mathcal{R}_i^1\} \cup \{\mathcal{R}_i^2\}$

Dekomposition in 4 NF



Zusammenfassung

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



abhängigkeitserh.
Zerlegung



verlustlose
Zerlegung

