

Kapitel 6

Relationale Entwurfstheorie

Funktionale Abhängigkeiten

Normalformen

Synthesealgorithmus

Dekomposition

Mehrwertige Abhängigkeiten

Ziele der relationalen Entwurfstheorie

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

Funktionale Abhängigkeiten

- Schema
 - $R = \{A, B, C, D\}$
- Ausprägung R

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

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:

$$CD \rightarrow B$$

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

- $\alpha \subseteq R$ ist ein **Super-Schlüssel**, falls folgendes gilt:
 - $\alpha \rightarrow R$
- 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: $\alpha \rightarrow \cdot \beta$
- $\alpha \subseteq R$ ist ein **Kandidaten-Schlüssel**, falls folgendes gilt:
 - $\alpha \rightarrow \cdot R$
- Ist $\alpha \subseteq R$ Kandidaten-Schlüssel, so nennt man alle $A \in \alpha$ **Schlüsselattribute**.
- **Nicht-Schlüsselattribute** heißen alle Attribute, die in *keinem* Kandidaten-Schlüssel vorkommen.

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
...

- Superschlüssel: {Kind,Vater,Mutter,Opa,Oma}, {Kind,Vater,Opa}, {Kind,Opa},
- Kandidatenschlüssel: {Kind,Opa}, {Kind,Oma}
- Schlüsselattribute: {Kind,Opa, Oma}
- Nicht-Schlüsselattribute: {Vater, Mutter}

Schlüsselbestimmung

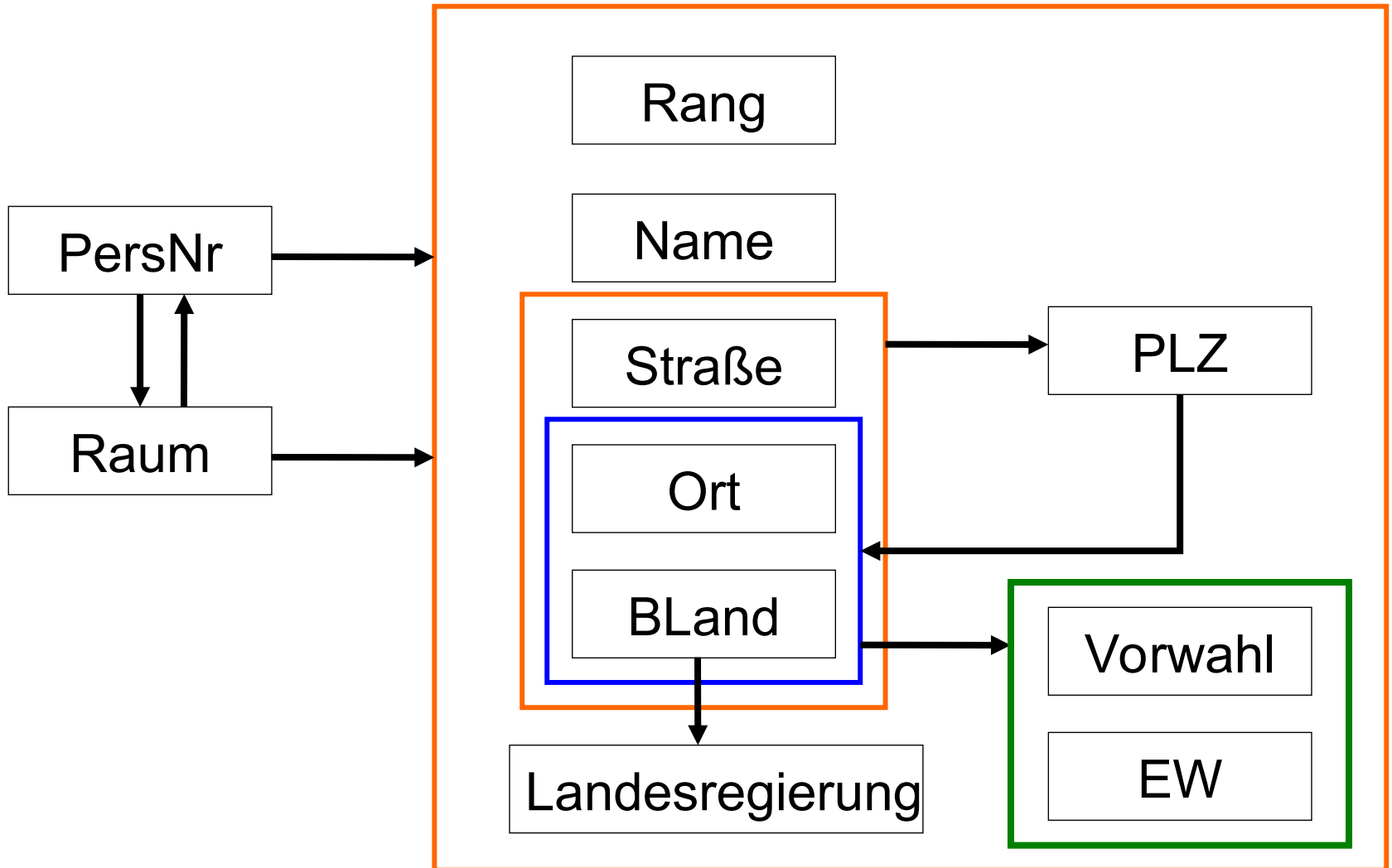
Städte			
Name	BLand	Vorwahl	EW
Frankfurt	Hessen	069	650000
Frankfurt	Brandenburg	0335	84000
München	Bayern	089	1200000
Passau	Bayern	0851	50000
...

- Funktionale Abhängigkeiten:
 - $\{\text{Name, BLand}\} \rightarrow \{\text{EW, Vorwahl}\}$
 - $\{\text{Name, Vorwahl}\} \rightarrow \{\text{BLand, EW}\}$
- Kandidaten-Schlüssel von *Städte*:
 - $\{\text{Name, BLand}\}$
 - $\{\text{Name, Vorwahl}\}$
- Beachte, dass 2 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



Bestimmung funktionaler Abhängigkeiten (zusätzliche Übung)

Vorlesungsverzeichnis									
VorlNr	Titel	SWS	gelesenVon	Vtag	Vzeit	Vraum	Uetag	Uezeit	Ueraum
001	Grundl. Inf.	4	007	Di	8	FMI 061	Do	12	FMI 003
001	Grundl. Inf.	4	007	Do	10	FMI 063	Fr	10	FMI 062
001	Grundl. Inf.	4	007	Do	10	FMI 063	Fr	14	FMI 003
...

- es gibt keine Parallelvorlesungen; aber es gibt u.U. mehrere Vorlesungen pro Woche (vielleicht auch am selben Tag)
 - es gibt mehrere Übungsgruppen; aber eine Übungsgruppe trifft sich nur einmal pro Woche
 - $\{\text{VorlNr}\} \rightarrow \{\text{Titel, SWS, gelesenVon}\}$
 - $\{\text{Vtag, Vzeit, VorlNr}\} \rightarrow \{\text{Vraum}\}$
(Eine Vorlesung kann sich vielleicht auch mehrmals am gleichen Tag treffen.)
 - $\{\text{Vtag, Vzeit, Vraum}\} \rightarrow \{\text{VorlNr}\}$
 - $\{\text{Uetag, Uezeit, Ueraum}\} \rightarrow \{\text{VorlNr}\}$
 - Schlüssel: $\{\text{Uetag, Uezeit, Ueraum, Vtag, Vzeit}\}$
- Weitere FD:
 $\{\text{gelesenVon, Vtag, Vzeit}\} \rightarrow \{\text{VorlNr}\}$

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 $\alpha\gamma \rightarrow \delta$

Beispiele zu Inferenzregeln

- Reflexivität: $\text{Name Semester} \rightarrow \text{Name}$
- Augmentation: $\text{Persnr} \rightarrow \text{Raum} \Rightarrow$
 $\text{Persnr Ort} \rightarrow \text{Raum Ort}$
- Transitivität: $\text{Vorlnr} \rightarrow \text{Persnr}, \text{Persnr} \rightarrow \text{Raum} \Rightarrow$
 $\text{Vorlnr} \rightarrow \text{Raum}$
- Vereinigung: $\text{Matrnr} \rightarrow \text{Name}, \text{Matrnr} \rightarrow \text{Semester} \Rightarrow$
 $\text{Matrnr} \rightarrow \text{Name Semester}$
- Dekomposition: $\text{Matrnr} \rightarrow \text{Name Semester} \Rightarrow$
 $\text{Matrnr} \rightarrow \text{Name}, \text{Matrnr} \rightarrow \text{Semester}$
- Pseudotransitivität: $\text{EAN} \rightarrow \text{Artikel}, \text{Artikel Menge} \rightarrow$
 $\text{Preis} \Rightarrow \text{EAN Menge} \rightarrow \text{Preis}$

Bestimmung der Hülle einer Attributmeng

- 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** $\beta \subseteq \text{Erg}$ **then** Erg := Erg \cup γ
 - Ausgabe $\alpha^+ = \text{Erg}$

Beispiel zur Hülle einer Attributmengenge

{PLZ} → {Bland, Ort}

{Ort, Bland} → {EW, Vorwahl}

{Bland, Ort, Straße} → {PLZ}

{Bland} → {Landesregierung}

- PLZ
- PLZ Bland Ort
- PLZ Bland Ort EW Vorwahl
- PLZ Bland Ort EW Vorwahl Landesregierung

Kanonische Überdeckung (minimale Menge von FDs)

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:

- $\forall A \in \alpha: (F_c - (\alpha \rightarrow \beta) \cup ((\alpha - \{A\}) \rightarrow \beta)) \not\equiv F_c$
- $\forall B \in \beta: (F_c - (\alpha \rightarrow \beta) \cup (\alpha \rightarrow (\beta - \{B\}))) \not\equiv F_c$

5) 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

- Ersetze alle $X \rightarrow \{A_1, A_2, \dots, A_n\}$ durch $X \rightarrow A_1, X \rightarrow A_2, \dots, X \rightarrow A_n$
- 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 $\beta \subseteq \text{AttrHülle}(F, \alpha - A)$ gilt.
Falls dies der Fall ist, ersetze $\alpha \rightarrow \beta$ durch $(\alpha - A) \rightarrow \beta$.
- Für jede FD $\alpha \rightarrow \beta \in F$:
 - Wenn $(F - \{\alpha \rightarrow \beta\})$ äquivalent zu F :
entferne $\alpha \rightarrow \beta$ aus F
- Fasse mittels der Vereinigungsregel 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.

Minimale Mengen von FDs (3)

Algorithmus nach Elmasri/Navathe mit Zusatz!

$AB \rightarrow CDE$

$B \rightarrow E$

$CD \rightarrow F$

Minimale Mengen von FDs (3)

$AB \rightarrow CDE$

$B \rightarrow E$

$CD \rightarrow F$

Jetzt: rechte Seite aufteilen

Minimale Mengen von FDs (3)

$AB \rightarrow C$

$AB \rightarrow D$

$AB \rightarrow E$

$B \rightarrow E$

$CD \rightarrow F$

Nun: Linke Seiten vereinfachen

Minimale Mengen von FDs (3)

$AB \rightarrow C$

~~$AB \rightarrow C$~~

$AB \rightarrow D$

$AB \rightarrow D$

$AB \rightarrow E$

$AB \rightarrow E$

$B \rightarrow E$

$B \rightarrow E$

$CD \rightarrow F$

$CD \rightarrow F$

Minimale Mengen von FDs (3)

$AB \rightarrow C$

~~$AB \rightarrow C$~~

$AB \rightarrow D$

$AB \rightarrow D$

$AB \rightarrow E$

$AB \rightarrow E$

$B \rightarrow E$

$B \rightarrow E$

$CD \rightarrow F$

$CD \rightarrow F$

Minimale Mengen von FDs (3)

$AB \rightarrow C$

$AB \rightarrow C$

$AB \rightarrow D$

~~$AB \rightarrow D$~~

$AB \rightarrow E$

$AB \rightarrow E$

$B \rightarrow E$

$B \rightarrow E$

$CD \rightarrow F$

$CD \rightarrow F$

Minimale Mengen von FDs (3)

$AB \rightarrow C$

$AB \rightarrow C$

$AB \rightarrow D$

~~$AB \rightarrow D$~~

$AB \rightarrow E$

$AB \rightarrow E$

$B \rightarrow E$

$B \rightarrow E$

$CD \rightarrow F$

$CD \rightarrow F$

Minimale Mengen von FDs (3)

$AB \rightarrow C$

$AB \rightarrow C$

$AB \rightarrow D$

$AB \rightarrow D$

$AB \rightarrow E$

~~$AB \rightarrow E$~~

$B \rightarrow E$

$B \rightarrow E$

$CD \rightarrow F$

$CD \rightarrow F$

Minimale Mengen von FDs (3)

$AB \rightarrow C$

$AB \rightarrow C$

$AB \rightarrow D$

$AB \rightarrow D$

$B \rightarrow E$

$B \rightarrow E$

$B \rightarrow E$

$B \rightarrow E$

$CD \rightarrow F$

~~$CD \rightarrow F$~~

Minimale Mengen von FDs (3)

$AB \rightarrow C$

$AB \rightarrow C$

$AB \rightarrow D$

$AB \rightarrow D$

$B \rightarrow E$

$B \rightarrow E$

$B \rightarrow E$

$B \rightarrow E$

$CD \rightarrow F$

~~$CD \rightarrow F$~~

Minimale Mengen von FDs (3)

$AB \rightarrow C$

$AB \rightarrow D$

$B \rightarrow E$

$B \rightarrow E$

$CD \rightarrow F$

Jetzt überflüssige Regeln streichen

Minimale Mengen von FDs (3)

$AB \rightarrow C$

$AB \rightarrow D$

$B \rightarrow E$

$CD \rightarrow F$

Zum Schluss: Regeln mit gleichen linken Seiten zusammenfassen

Minimale Mengen von FDs (3)

$AB \rightarrow CD$

$B \rightarrow E$

$CD \rightarrow F$

Fertig!

„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

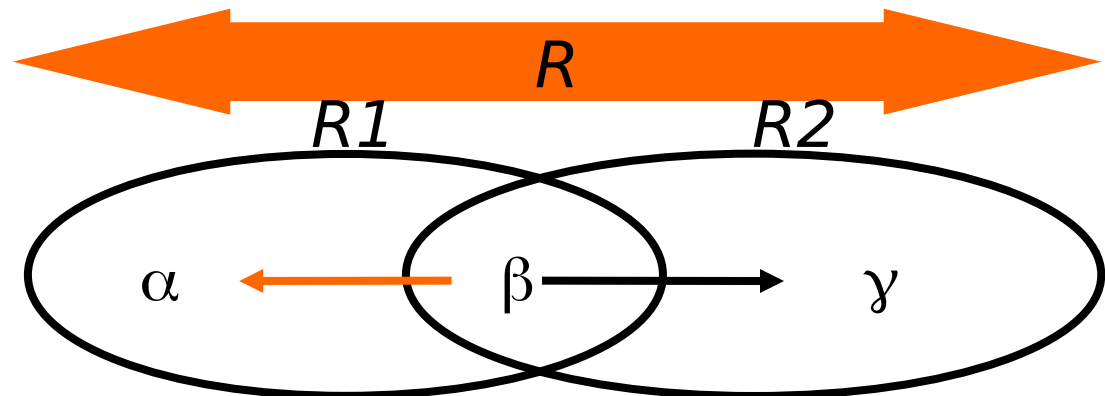
- Die in der ursprünglichen Relationenausprägung R des Schemas R enthaltenen Informationen müssen aus den Ausprägungen R_1, \dots, R_n der neuen Relationenschemata R_1, \dots, R_n rekonstruierbar sein.

2. Abhängigkeitserhaltung

- Die für R geltenden funktionalen Abhängigkeiten müssen auf die Schemata R_1, \dots, R_n übertragbar sein.

Kriterien für die Verlustlosigkeit einer Zerlegung

- $R = R1 \cup R2$
 - $R1 := \Pi_{R1} (R)$
 - $R2 := \Pi_{R2} (R)$
- Die Zerlegung von R in $R1$ und $R2$ ist verlustlos, falls für *jede mögliche (gültige) Ausprägung* R von R gilt:
 - $R = R1 \bowtie R2$
- Hinreichende Bedingung für die Verlustlosigkeit einer Zerlegung
 - $(R1 \cap R2) \rightarrow R1$ oder
 - $(R1 \cap R2) \rightarrow R2$



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

Π _{Kneipe, Gast}

Π _{Gast, Bier}

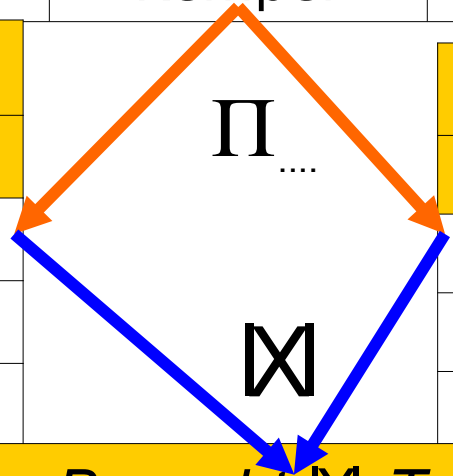
<i>Besucht</i>	
<i>Kneipe</i>	<i>Gast</i>
Kowalski	Kemper
Kowalski	Eickler
Innsteg	Kemper

<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



<i>Besucht</i> \bowtie <i>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

≠



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

- R ist zerlegt in R_1, \dots, R_n
- $F_R = (F_{R_1} \cup \dots \cup F_{R_n})$ bzw. $F_{R^+} = (F_{R_1} \cup \dots \cup F_{R_n})^+$
- 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

$\Pi_{PLZ, Straße}$

$\Pi_{Stadt, BLand, PLZ}$

Straßen

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

Orte

<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

$\Pi_{PLZ, Straße}$

$\Pi_{Stadt, Bland, PLZ}$

Straßen

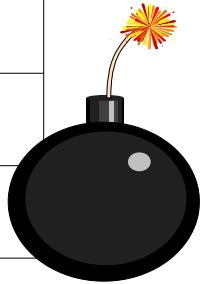
Orte

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

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



X

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

<i>Orte</i>		
Ort	BLand	PLZ
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

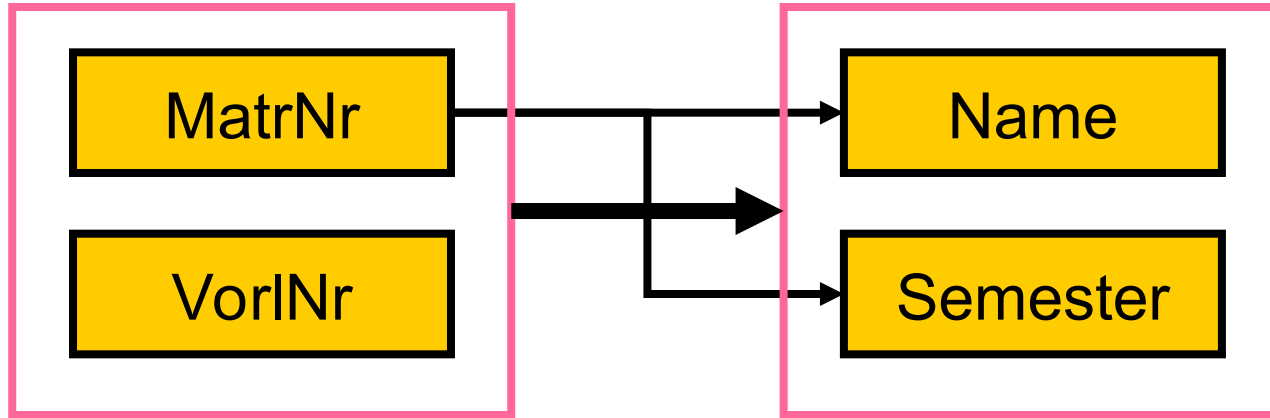
Zweite Normalform

- Eine Relation R mit zugehörigen FDs F_R ist in zweiter Normalform, falls jedes Nichtschlüssel-Attribut $A \in R$ 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

- Ein Relationenschema R ist in dritter Normalform, wenn für jede für R geltende funktionale Abhängigkeit der Form $\alpha \rightarrow B$ mit $B \in R$ und mindestens **eine** von drei Bedingungen gilt:
 - $B \in \alpha$, d.h., die FD ist trivial
 - Das Attribut B ist ein Schlüsselattribut
 - α ist Superschlüssel von R
- **Alternative Formulierung (einfacher):**
Für alle Nicht-Schlüsselattribute A gilt:
für alle α mit $\alpha \rightarrow A$ ist α Kandidatenschlüssel
(A ist nicht transitiv abhängig von einem Kandidatenschlüssel).

Beispiele zu 3NF

- Studenten: {[MatrNr, Name, Semester]}
MatrNr → Name Semester
 - in 3NF, da linke Seite Kandidatenschlüssel ✓
- Lieferung(KundenNr, Adresse, Entfernung)
KundenNr -> Adresse,
Adresse -> Entfernung
 - in 2NF, da alle NSA vom ganzen Schlüssel abh. ✓
 - nicht in 3NF, da Entfernung von einem NSA abh. ✗

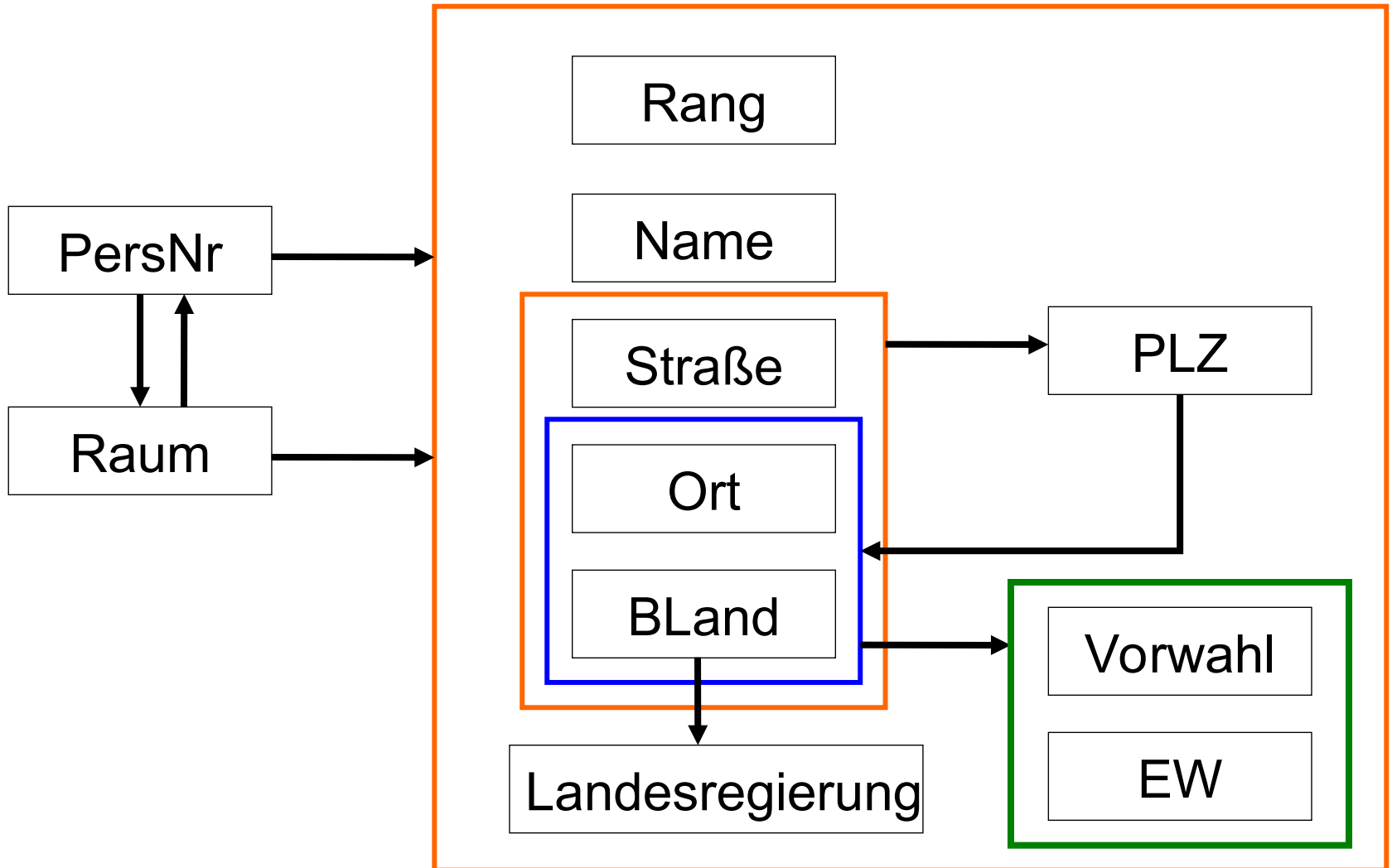
Zerlegung mit dem Synthesealgorithmus

- Wir geben jetzt einen sogenannten Synthesealgorithmus an, mit dem zu einem gegebenen Relationenschema R mit funktionalen Abhängigkeiten F eine Zerlegung in R_1, \dots, R_n ermittelt wird, die alle drei folgenden Kriterien erfüllt.
 - R_1, \dots, R_n ist eine verlustlose Zerlegung von R .
 - Die Zerlegung R_1, \dots, R_n ist abhängigkeiterhaltend.
 - Alle R_1, \dots, R_n sind in dritter Normalform.

Synthesealgorithmus

- Bestimme die kanonische Überdeckung F_c zu F .
Wiederholung:
 - einelementige rechte Seiten
 - Linksreduktion
 - Elimination redundanter funktionaler Abhängigkeiten
 - Zusammenfassung gleicher linker Seiten
- Für jede funktionale Abhängigkeit $\alpha \rightarrow \beta \in F_c$:
 - Kreiere ein Relationenschema $R_\alpha := \alpha \cup \beta$
 - Ordne R_α die FDs $F_\alpha := \{\alpha' \rightarrow \beta' \in F_c \mid \alpha' \cup \beta' \subseteq R_\alpha\}$ zu.
- Falls eines der in Schritt 2. erzeugten Schemata einen Kandidatenschlüssel von R bzgl. F_c enthält, sind wir fertig.
Sonst wähle einen Kandidatenschlüssel $\kappa \subseteq R$ aus und definiere folgendes Schema:
 - $R_\kappa := \kappa$
 - $F_\kappa := \emptyset$
- Eliminiere diejenigen Schemata R_α , die in einem anderen Relationenschema $R_{\alpha'}$ enthalten sind, d.h.,
 - $R_\alpha \subseteq R_{\alpha'}$

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)

● ~~RP(Raum, PersNr)~~

● PLZV(Straße, BLand, Ort, PLZ)

● OrtsV(Ort, Bland, EW, Vorwahl)

● Regierungen(BLand, Landesregierung)

● ~~PV(PLZ, Bland, Ort)~~

StudVorl(MatrNr, Name, Semester, VorlNr, Titel, SWS)

MatrNr → Name Semester

VorlNr → Titel SWS



- Studenten(MatrNr, Name, Semester)
- Vorlesungen(VorlNr, Titel, SWS)
- Hört(MatrNr, VorlNr)

Boyce-Codd-Normalform

- Die Boyce-Codd-Normalform (BCNF) ist nochmals eine Verschärfung der 3 NF.
- Ein Relationenschema R mit FDs F ist in BCNF, wenn für jede für R geltende funktionale Abhängigkeit der Form $\alpha \rightarrow \beta \in F$ und mindestens **eine** von zwei Bedingungen gilt:
 - $\beta \subseteq \alpha$, d.h., die Abhängigkeit ist trivial oder
 - α ist Superschlüssel von R
- **Alternative Formulierung:** *Alle Attribute A hängen vollständig nur von Kandidatenschlüsseln ab:*
für alle α mit $\alpha \rightarrow \cdot A$ gilt: α ist Kandidatenschlüssel
- Man kann jede Relation **verlustlos** in BCNF-Relationen zerlegen
- Manchmal lässt sich dabei die **Abhängigkeitserhaltung** aber **nicht** erzielen

bei 3NF nur die NSA!

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}

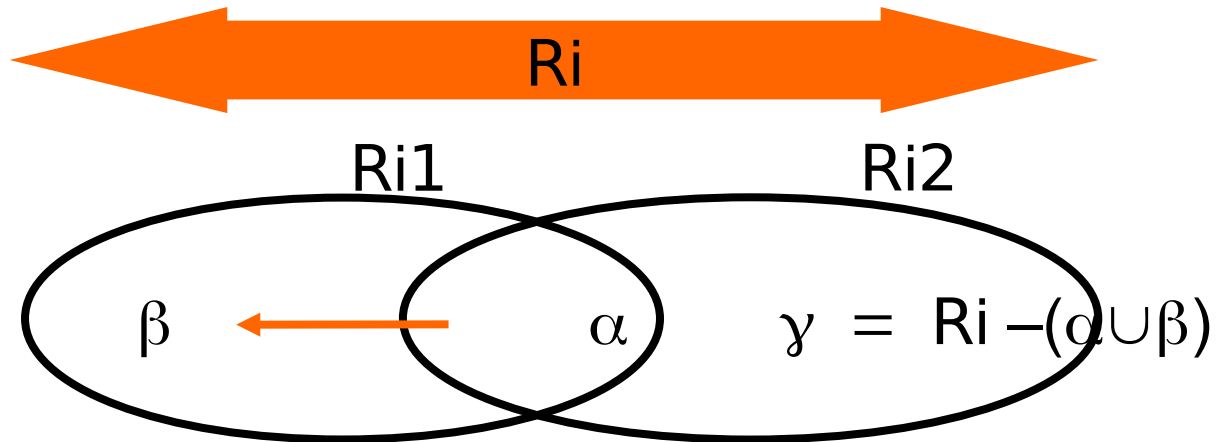
Dekomposition

- Man kann grundsätzlich jedes Relationenschema 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 leider nicht immer erreicht werden, dass die Zerlegung R_1, \dots, R_n abhängigkeiterhaltend ist.


Dekompositions-Algorithmus

- Starte mit $Z = \{R\}$
- Solange es noch ein Relationenschema R_i in Z gibt, das nicht in BCNF ist, mache folgendes:
 - Es gibt also eine für R_i geltende nicht-triviale funktionale Abhängigkeit $(\alpha \rightarrow \beta)$ mit
 - $\alpha \cap \beta = \emptyset$
 - $\neg(\alpha \rightarrow R_i)$
 - 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.
 - 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
 - $Z := (Z - \{R_i\}) \cup \{R_{i1}\} \cup \{R_{i2}\}$

Veranschaulichung der Dekomposition

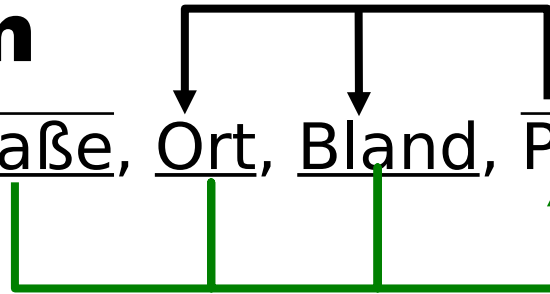


Dekomposition der Relation Städte in BCNF-Relationen

- Städte: $\{\overline{[Ort, BLand, Ministerpräsident/in, EW]}\}$
- Geltende FDs:
 - $\{BLand\} \rightarrow \{Ministerpräsident/in\}$ 
 - $\{Ort, BLand\} \rightarrow \{EW\}$
 - $\{Ministerpräsident/in\} \rightarrow \{BLand\}$
- Ri1:
 - Regierungen: $\{[BLand, Ministerpräsident/in]\}$
- Ri2:
 - Städte: $\{[Ort, BLand, EW]\}$
- Zerlegung ist verlustlos und auch abhängigkeiterhaltend

Dekomposition des PLZverzeichnis in BCNF-Relationen

- PLZverzeichnis: $\{\overline{[Stra\ddot{u}\beta e, Ort, BLand, PLZ]}\}$



- Funktionale Abhängigkeiten:
 - $\{PLZ\} \rightarrow \{Ort, BLand\}$ ⚡
 - $\{Stra\ddot{u}\beta e, Ort, BLand\} \rightarrow \{PLZ\}$
- Betrachte die Zerlegung
 - Orte: $\{[PLZ, Ort, BLand]\}$
 - Straßen: $\{[PLZ, Stra\ddot{u}\beta e]\}$
- Diese Zerlegung
 - ist verlustlos
 - aber nicht abhängigkeiterhaltend
 - siehe oben

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 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
3005	deutsch

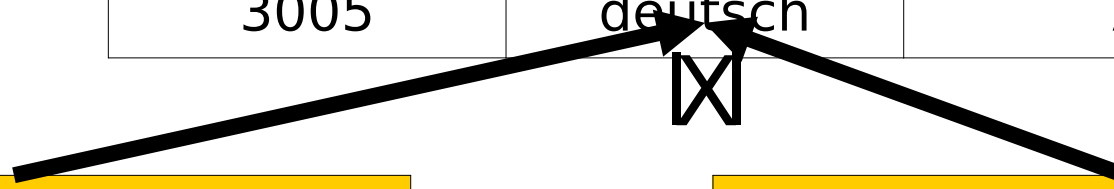
PSprachen	
PersNr	ProgSprache
3002	C
3002	Pascal
3005	Ada

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
3005	deutsch

PSprachen	
PersNr	ProgSprache
3002	C
3002	Pascal
3005	Ada



Mehrwertige Abhängigkeiten

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

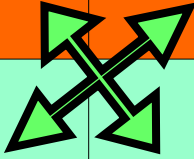
- $\alpha \twoheadrightarrow \beta$ gilt genau dann wenn
 - Wenn es zwei Tupel t_1 und t_2 mit gleichen α -Werten gibt
 - Dann muss es auch zwei Tupel t_3 und t_4 geben mit
 - $t_3.\alpha = t_4.\alpha = t_1.\alpha = t_2.\alpha$
 - $t_3.\beta = t_2.\beta$, $t_4.\beta = t_1.\beta$
 - $t_3.\gamma = t_1.\gamma$, $t_4.\gamma = t_2.\gamma$

MVDs

- 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



● $A \twoheadrightarrow B$

● $A \twoheadrightarrow C$

Verlustlose Zerlegung bei MVDs: hinreichende + notwendige Bedingung

- $R = R1 \cup R2$
 - $R1 := \Pi_{R1}(R)$
 - $R2 := \Pi_{R2}(R)$
- Die Zerlegung von R in $R1$ und $R2$ ist verlustlos, falls für *jede mögliche (gültige) Ausprägung* R von R gilt:
 - $R = R1 \bowtie R2$
- Die Zerlegung von R in $R1$ und $R2$ ist verlustlos **genau dann wenn**
 - $R = R1 \cup R2$
 - und** mindestens eine von zwei MVDs gilt:
 - $(R1 \cap R2) \twoheadrightarrow R1$ oder
 - $(R1 \cap R2) \twoheadrightarrow R2$

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 ...

- ... sind solche, die von jeder Relationenausprägung erfüllt werden
- Eine MVD $\alpha \twoheadrightarrow \beta$ ist trivial genau dann wenn
 - $\beta \subseteq \alpha$ oder
 - $\beta = R - \alpha$

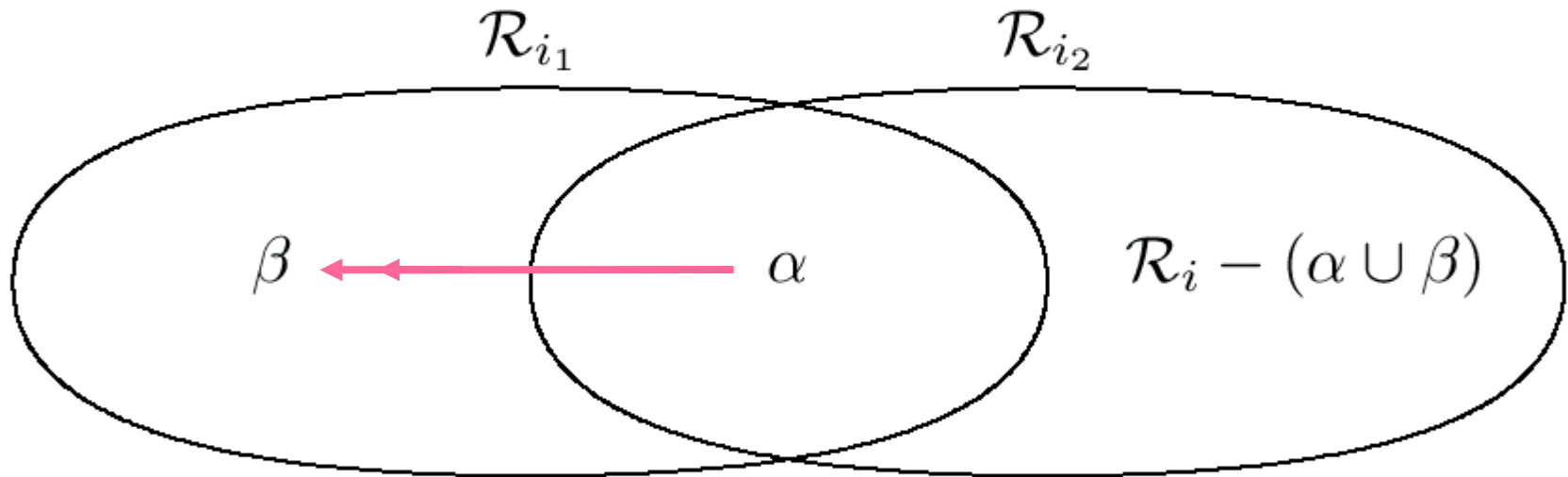
Vierte Normalform

- Eine Relation 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 R

Dekomposition in 4 NF

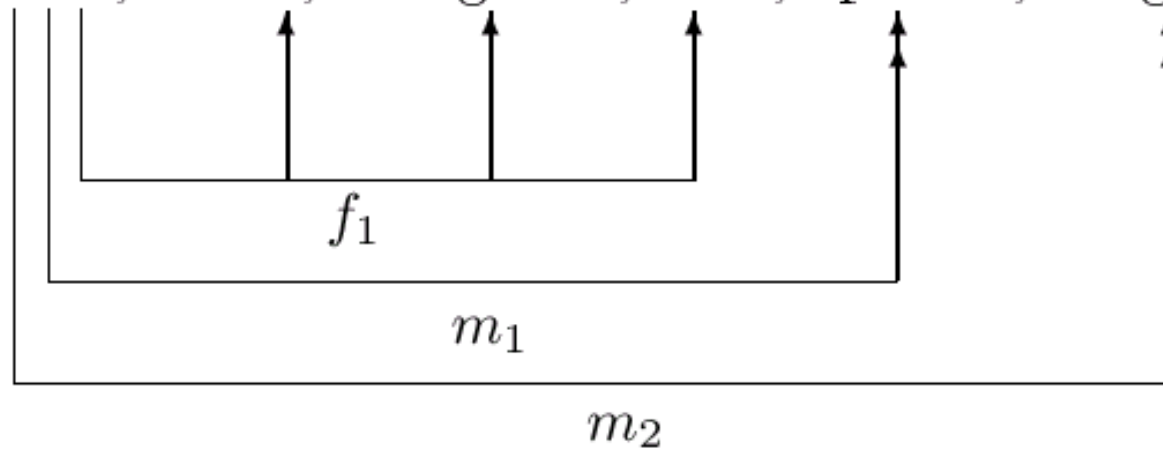
- Starte mit der Menge $Z := \{R\}$
- Solange es noch ein Relationenschema R_i in Z gibt, das nicht in 4NF ist, mache folgendes:
 - Es gibt also eine für R_i geltende nicht-triviale MVD $(\alpha \twoheadrightarrow \beta)$, für die gilt:
 - $\alpha \cap \beta = \emptyset$
 - $\neg(\alpha \rightarrow R_i)$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

Dekomposition in 4 NF



Beispiel-Zerlegung

Assistenten': {[PersNr, Name, Fachgebiet, Boss, Sprache, ProgrSprache]}



- Assistenten: {[PersNr, Name, Fachgebiet, Boss]}
- Fähigkeiten: {[PersNr, Sprache, ProgrSprache]}
- Sprachen: {[PersNr, Sprache]}
- ProgrSprachen: {[PersNr, ProgrSprache]}

Beispiel-Zerlegung (2)

Assistenten'(PersNr, Name, Fachgebiet, Boss, Sprache, ProgSprache)

PersNr → Name Fachgebiet Boss

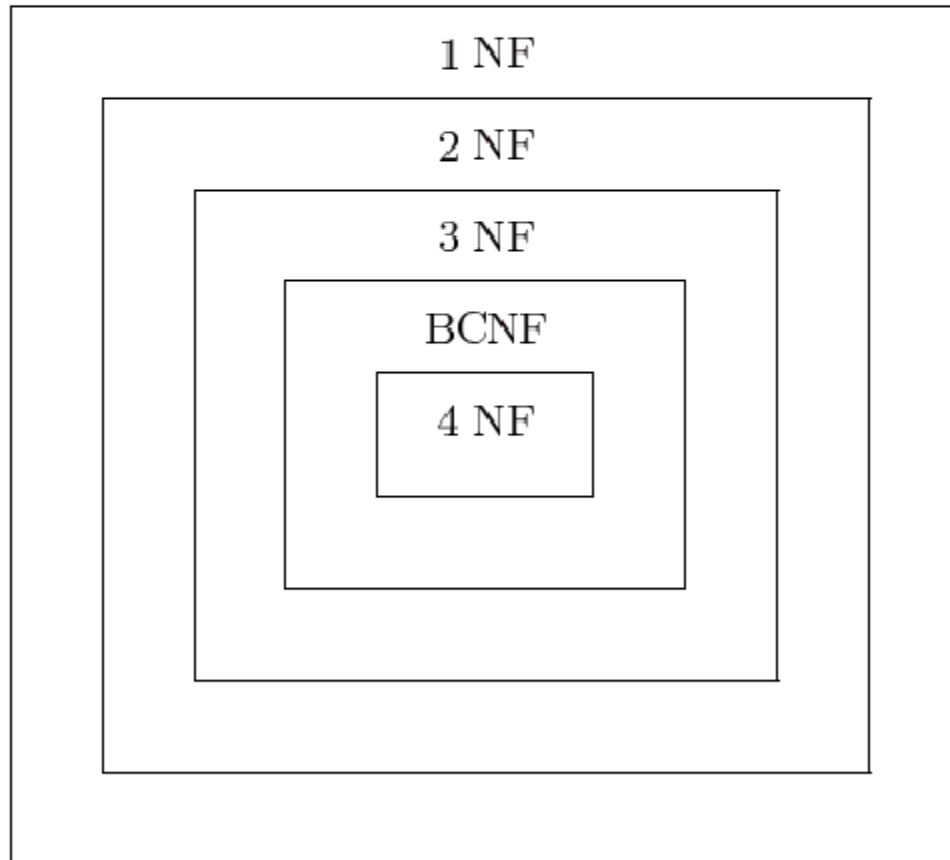
PersNr ->> Sprache

PersNr ->> ProgSprache

- Assistenten(PersNr, Name, Fachgebiet, Boss)
- ~~ASP(PersNr, Sprache, ProgSprache)~~
- Sprachen(PersNr, Sprache)
- ProgSp(PersNr, ProgSprache)

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



Übung: FDs, MVDs, Normalisierung

Vorlesungen: {[VorlNr, Titel, SWS, gelesenVon, VTermin, VRaum, ÜTermin, ÜRaum]}

- FDs
 - VorlNr \rightarrow Titel, SWS, gelesenVon
 - VRaum, VTermin \rightarrow VorlNr
 - VorlNr, VTermin \rightarrow VRaum
 - ÜTermin, ÜRaum \rightarrow VorlNr
- Schlüssel: {VTermin, ÜTermin, ÜRaum}
- MVDs
 - VorlNr \twoheadrightarrow VTermin, VRaum
 - VorlNr \twoheadrightarrow ÜTermin, ÜRaum

Normalisierung: 3NF, BCNF, 4NF

Übung(2)

Vorlesungen(VorlNr, Titel, SWS, gelesen Von, VTermin, VRaum, ÜTermin, ÜRaum)

- VorlNr → Titel SWS gelesen Von
 - Vorlesungen(VorlNr, Titel, SWS, gelesen Von) 4NF
 - R1(VorlNr, VTermin, VRaum, ÜTermin, ÜRaum) -2NF
- VRaum Vtermin → VorlNr
 - R11(Vraum, Vtermin, VorlNr) 4NF
 - R12(VTermin, VRaum, Ütermin, Üraum) 4NF, nicht abh.-erhaltend
- alternative Zerlegung von R1:
- VorlNr ->> Vtermin VRaum
 - R11'(VorlNr, Vtermin, VRaum) 4NF
 - R12'(VorlNr, ÜTermin, ÜRaum) 4NF, abh.-erhaltend

Weitere Übung: Familie

- 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
- Abhängigkeiten:
 - $K \rightarrow V, M$
 - $K, Opa \rightarrow Oma$
 - $K, Oma \rightarrow Opa$
 - $V, M \rightarrow\rightarrow K$
 - $V, M \rightarrow\rightarrow Opa, Oma$

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
Tobias	Leo	Bertha	Hubert	Martha
...

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

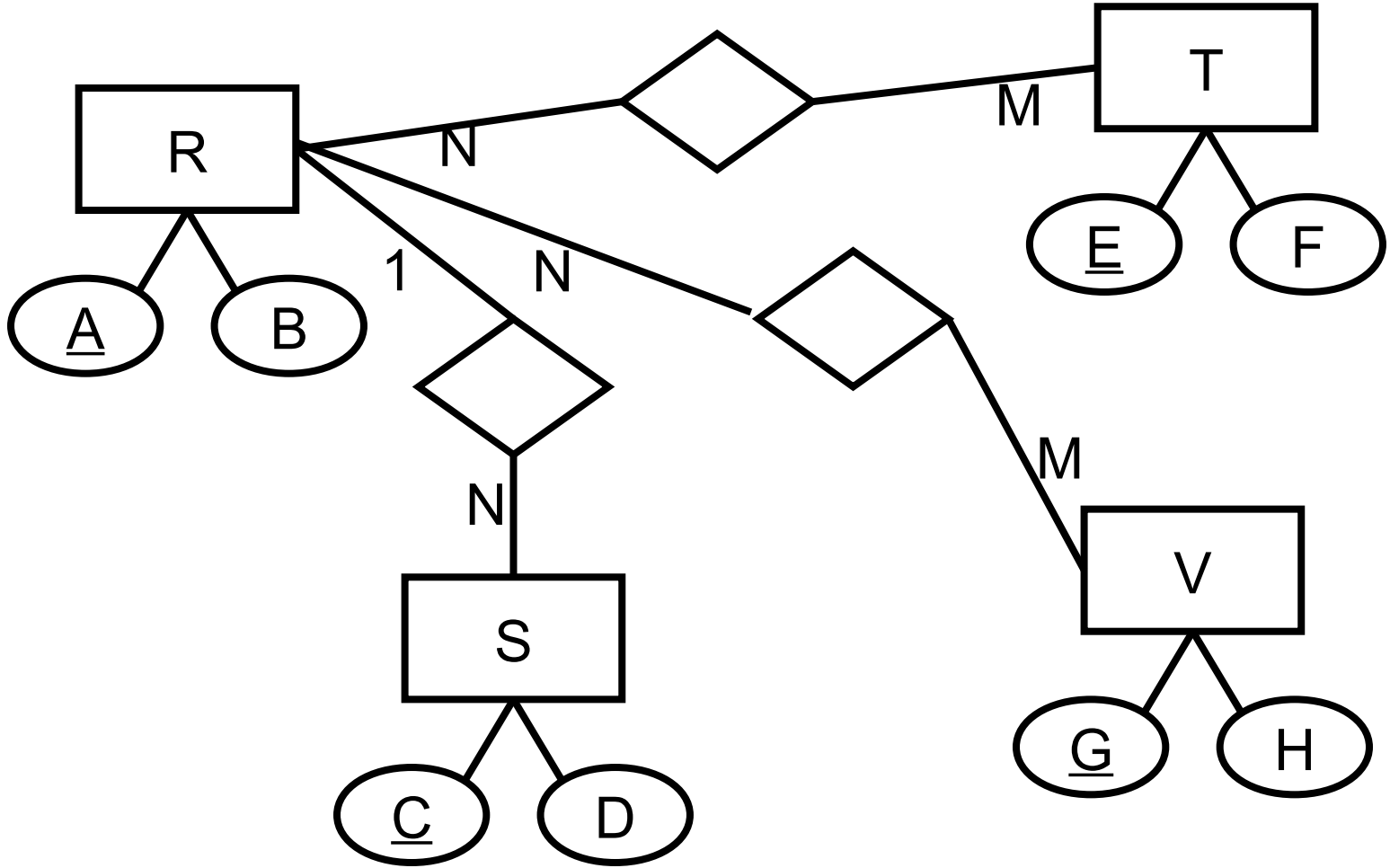
Zerlegung

Familie(Opa, Oma, Vater, Mutter, Kind)

Kind → Vater Mutter

- Eltern(Kind, Vater, Mutter) 4NF
 - Großeltern(Kind, Opa, Oma) 4NF, nicht abh.-erhaltend
- alternative Zerlegung:
 - Vater Mutter ->> Kind
 - Eltern(Kind, Vater, Mutter) 4NF
 - EG(Vater, Mutter, Opa, Oma) 4NF, nicht abh.-erhaltend

Weiteres Beispiel: 1:N & N:M Bez.



UR: { [A , B , C , D , E , F , G , H] }

UR(A, B, C, D, E, F, G, H)

A → B

● R(A,B)

C → D

● UR'(A, C, D, E, F, G, H)

E → F

● S(C,D)

G → H

● UR''(A, C, E, F, G, H)

C → A

● T(E,F)

A-→>C

● UR'''(A, C, E, G, H)

A-→>E

● V(G,H)

E-→>A

● UR''''(A, C, E, G)

A-→>G

● U(C,A)

G-→>A

● UR'''''(C, E, G)

E-→>G

● W(E,G)

C-→>G

● X(C,E)

C-→>E