Grundlagen der Datenbanken

Relationale Entwurfstheorie

Dr. Jérôme Kunegis

Wintersemester 2012/13

WeST
People and Knowledge Networks

Lernziele

- Charakterisierung "guter" relationaler Schemata:

  - jede Relation entspricht genau einer Objektmenge - eventuell unter Einbezug von N:1- oder 1:1-Relationships - oder genau einer Relationship-Menge zwischen Objekten

  - Redundanz ist eliminiert, alle Informationen sind repräsentierbar, und es treten keinerlei "Änderungsanomalien" auf

    a) Änderungen können bei Beachtung der Primärschlüssel- und Fremdschlüsselbedingung keine Inkonsistenzen hervorrufen

    b) alle Informationen lassen sich unter Wahrung der Primärschlüssel- und Fremdschlüsselbedingung (ohne "Kunstgriffe") einfügen

    c) Informationen können einzeln wieder gelöscht werden, ohne die Primärschlüssel oder Fremdschlüsselbedingung zu verletzen

- Algorithmische Herleitung solcher "guter" Schemata

Beispiel: "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	Schlick	C3	52	5259	Der Wiener Kreis	2
2137	Kant	C4	7	4630	Die 3 Kritiken	4

Beispiel: "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	Schlick	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?

Gliederung

- Funktionale Abhängigkeiten
- Dekomposition der Relationenschemata: Verlustlosigkeit und Abhängigkeitsbewahrung
- Normalformen als Gütekriterium
- Erste Normalform
- Zweite Normalform
- Dritte Normalform, Zerlegung mit dem Synthesealgorithmus
- Boyce–Codd Normalform (BCNF), Dekompositions-Algorithmus

Relationenmodell revisited: Definitionen

Gegeben sei eine Menge von Wertebereichen primitiver Datentypen $\{D_1, ..., D_m\}$, die als "***Domains***" bezeichnet werden.

Eine ***Relation*** R ist ein Paar $R = (s,v)$ mit

- einem *Schema* $s = sch(R) = \{A_1, ..., A_n\}$, das aus einer Menge von ***Attributen*** (***Attributnamen***) besteht und für jedes Attribut A_i einen Domain $dom(A_i) \in \{D_1, ..., D_m\}$ festlegt, und

- einer ***Ausprägung***
  $v = val(R) \subseteq dom(A_1) \times dom(A_2) \times ... \times dom(A_n)$.

Funktionale Abhängigkeiten (FD)

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

Seien $\alpha \subseteq \boldsymbol{R}$, $\beta \subseteq \boldsymbol{R}$
$\alpha \to \beta$ (gesprochen: α bestimmt β) genau dann wenn

$$\forall r, s \in Q \ (r.\alpha = s.\alpha \Rightarrow r.\beta = s.\beta)$$

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

$\{A\} \to \{B\}$

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

Nicht: $\{B\} \to \{C\}$

Notationskonvention:

$CD \to B$

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
…	…	…	…	…

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 $\rightarrow$ Vater, Mutter

Kind, Opa $\rightarrow$ Oma

Kind, Oma $\rightarrow$ Opa

Schlüssel

$\alpha \subseteq \boldsymbol{R}$ ist ein **Super-Schlüssel**, falls folgendes gilt:

- $\alpha \rightarrow \boldsymbol{R}$

β ist **voll funktional abhängig** von α genau dann wenn gilt

- $\alpha \rightarrow \beta$ und
- α kann nicht mehr verkleinert werden, d.h.

  $$\forall A \in \alpha \;(\neg((\alpha - \{A\}) \rightarrow \beta))$$

Notation für volle funktionale Abhängigkeit: $\alpha \rightarrow^{\bullet} \beta$

$\alpha \subseteq \boldsymbol{R}$ ist ein **Kandidaten-Schlüssel**, falls folgendes gilt:

- $\alpha \rightarrow^{\bullet} \boldsymbol{R}$

Attribut $\boldsymbol{B}$ heißt **Schlüsselattribut**, wenn $\boldsymbol{B}$ in einem von Kandidaten-Schlüsseln $\alpha \subseteq \boldsymbol{R}$ vorkommt

Schlüsselbestimmung

Städte			
Name	Bundesland	Vorwahl	Einwohner
Frankfurt	Hessen	069	650.000
Frankfurt	Brandenburg	0335	84.000
München	Bayern	089	1.200.000
Mülheim-Kärlich	RLP	0261	10.416
Koblenz	RLP	0261	106.681

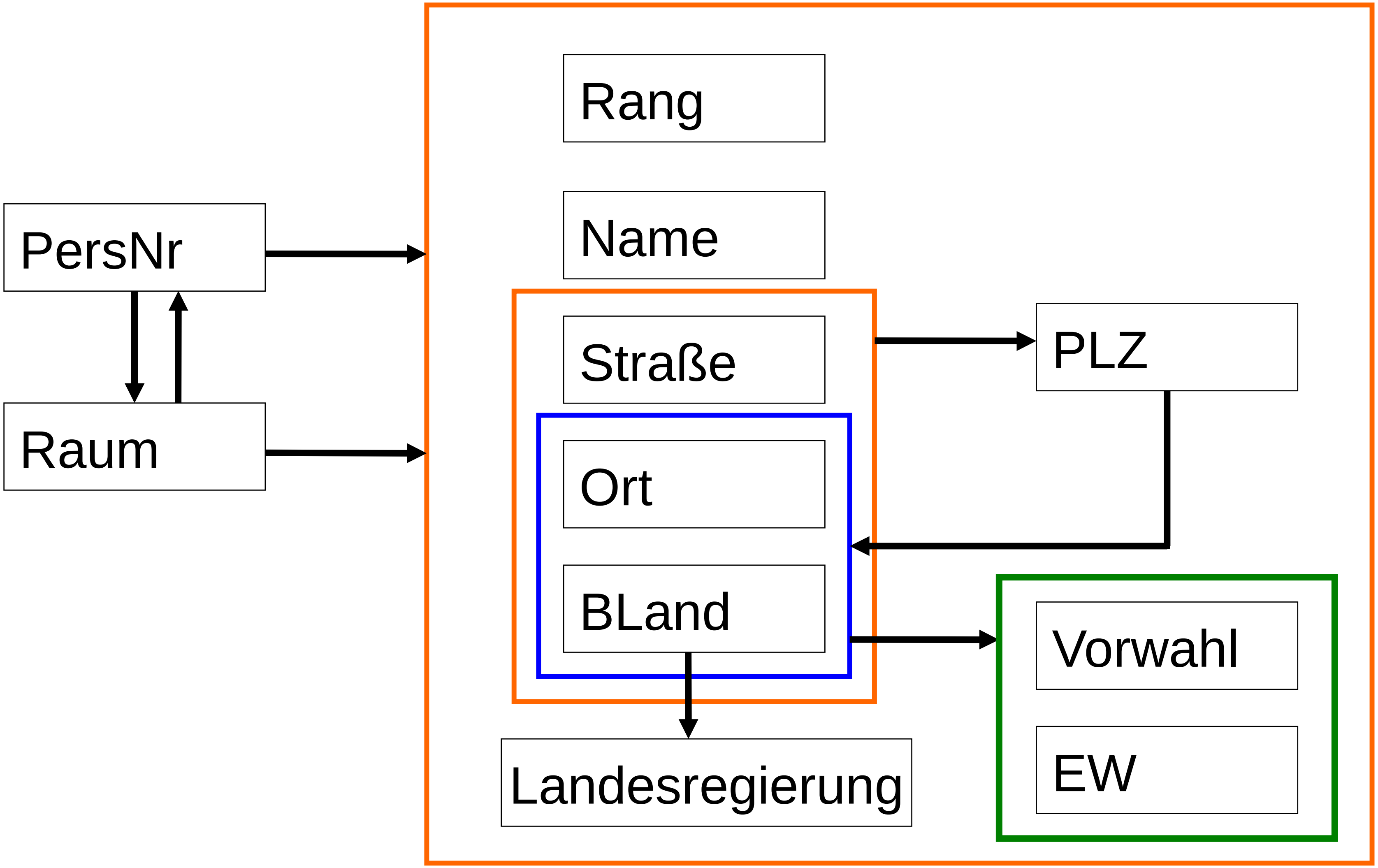
Kandidaten-schlüssel von *Städte*:

- {Name, Bundesland}
- {Name, Vorwahl}

Keine Kandidaten-Schlüssel:

- {Vorwahl}, {Name}, {Einwohner, Bundesland}

Graphische Darstellung der funktionalen Abhängigkeiten



Herleitung funktionaler Abhängigkeiten: Armstrong-Axiome

Reflexivität

- Falls β eine Teilmenge von α ist ($\beta \subseteq \alpha$) dann gilt immer $\alpha \to \beta$. Insbesondere gilt immer $\alpha \to \alpha$.

Verstärkung

- Falls $\alpha \to \beta$ gilt, dann gilt auch $\alpha\gamma \to \beta\gamma$. ($\alpha\gamma$ steht hier für $\alpha \cup \gamma$)

Transitivität

- Falls $\alpha \to \beta$ und $\beta \to \gamma$ gilt, dann gilt auch $\alpha \to \gamma$.

Diese drei Axiome sind vollständig und korrekt.
Zusätzliche Axiome erleichtern die Herleitung:

- Vereinigungsregel:
  - Wenn $\alpha \to \beta$ und $\alpha \to \gamma$ gelten, dann gilt auch $\alpha \to \beta\gamma$
- Dekompositionsregel:
  - Wenn $\alpha \to \beta\gamma$ gilt, dann gelten auch $\alpha \to \beta$ und $\alpha \to \gamma$
- Pseudotransitivitätsregel:
  - Wenn $\alpha \to \beta$ und $\gamma\beta \to \delta$, dann gilt auch $\alpha\gamma \to \delta$

Bestimmung der Hülle einer Attributmenge

Eingabe:

eine Menge F von funktionalen Abhängigkeiten (FD) und eine Menge von Attributen α.

Ausgabe:

die vollständige Menge von Attributen α^+, für die gilt $\alpha \rightarrow \alpha^+$.

AttrHülle ($\boldsymbol{F}$, α)

result := α

WHILE (Änderungen an result) **DO**

FOREACH FD $\beta \rightarrow \gamma$ **IN** F **DO**

IF $\beta \subseteq$ result **THEN** result := result $\cup$ γ

Ausgabe α^+ = result

Transitive Hülle der FDs

Gegeben sei eine Menge $\boldsymbol{F}$ von FDs

Die Menge aller aus $\boldsymbol{F}$ logisch ableitbaren FDs
wird als **transitive Hülle** $\boldsymbol{F}^{+}$ bezeichnet

Satz:

Sei R eine Relation mit einer Menge $\boldsymbol{F}$ von FDs.

Durch endlich-malige Anwendung von Armstrong-Regeln
lassen sich genau alle Funktionalabhängigkeiten in $\boldsymbol{F}^{+}$ herleiten

Die Armstrong-Regeln bilden also einen korrekten und vollständigen
Ableitungskalkül für die FDs in $\boldsymbol{F}^{+}$

Bestimmung funktionaler Abhängigkeiten

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

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

Zusätzliche Abhängigkeiten, die aus obigen abgeleitet werden können:

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

Kanonische Überdeckung

Fc heißt **kanonische Überdeckung** von F,
wenn die folgenden drei Kriterien erfüllt sind:

1. $Fc \equiv F \Leftrightarrow Fc^{+} = F^{+}$

1. In Fc existieren keine FDs $\alpha \rightarrow \beta$, die überflüssige Attribute enthalten:

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

1. Jede linke Seite einer funktionalen Abhängigkeit in Fc ist einzigartig.
   Erreichbar durch sukzessive Anwendung der Vereinigungsregel:
   $\alpha \rightarrow \beta$ und $\alpha \rightarrow \gamma$ zu $\alpha \rightarrow \beta\gamma$

Berechnung der kanonischen Überdeckung

1) Führe für jede FD $\alpha \rightarrow \beta \in \boldsymbol{F}$ die **Linksreduktion** durch:
   - Überprüfe für alle $A \in \alpha$, ob A überflüssig ist, d.h., ob

     $\beta \subseteq AttrHülle(F, \alpha - A)$ 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

     $B \in AttrHülle(F - (\alpha \rightarrow \beta) \cup (\alpha \rightarrow (\beta - B)), \alpha)$ 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 Schritt 2) möglicherweise entstanden sind.

4) **Fasse** mittels der Vereinigungsregel FDs der Form $\alpha \rightarrow \beta_1, \ldots, \alpha \rightarrow \beta_n$ **zusammen** zu $\alpha \rightarrow (\beta_1 \cup \ldots \cup \beta_n)$.

Beispiele: Links/Rechtsreduktion

(1)

$$AB \rightarrow C$$
$$B \rightarrow C$$

(2)

$$A \rightarrow BC$$
$$B \rightarrow C$$

(3)

$$A \rightarrow B$$
$$B \rightarrow C$$
$$A \rightarrow C$$

Zerlegung (Dekomposition) von Relationen

Korrektheitskriterien für die Zerlegung von Relationenschemata:

Verlustlosigkeit

- Die in der ursprünglichen Relationenausprägung $\boldsymbol{Q}$ des Schemas $\boldsymbol{R}$ enthaltenen Informationen müssen aus den Ausprägungen $\boldsymbol{Q_1}$, ..., $\boldsymbol{Q_n}$ der neuen Relationenschemata $\boldsymbol{R_1}$, .., $\boldsymbol{R_n}$ rekonstruierbar sein.

Abhängigkeitserhaltung

- Die für $\boldsymbol{R}$ geltenden funktionalen Anhängigkeiten müssen auf die Schemata $\boldsymbol{R_1}$, ..., $\boldsymbol{R_n}$ übertragbar sein.

Definition: Projektion

Projektion π (Auswahl von Spalten einer Tabelle):

Sei $A \subseteq sch(R)$. Das Resultat einer Projektion $\pi[A](R)$ ist:

$$sch(\pi[A](R)) = A$$
$$val(\sigma[F](R)) = \{t \mid \exists r \in val(R)(t.A = r.A)\}$$

Definition: Natural Join $\bowtie$

Natural Join:

Natürlicher Verbund von Relationen
über gleiche Attributnamen und Attributwerte.

Das Resultat von $R \bowtie S$ mit $A = sch(R)$ und $B = sch(S)$ ist:

$$\mathrm{sch}(R \bowtie S) = \mathrm{sch}(R) \cup \mathrm{sch}(S)$$
$$\mathrm{val}(R \bowtie S) = \{\ t \mid \exists\, r \in \mathrm{val}(R), s \in \mathrm{val}(S):\ t.A = r.A \wedge t.B = s.B\ \}$$

$$P\ (X_1, \ldots, X_m, Y_1, \ldots, Y_k)$$
$$Q\ (Y_1, \ldots, Y_k, Z_1, \ldots, Z_n)$$

$$P \bowtie Q$$

P – Q				P ∩ Q				Q – P			
X_1	X_2	...	X_m	Y_1	Y_2	...	Y_k	Z_1	Z_2	...	Z_n

22

Kriterien für die Verlustlosigkeit einer Zerlegung

$$Rel := (R, Q) \qquad R = R_1 \cup R_2$$

$$Rel_1 := (R_1, Q_1), \quad Q_1 := \pi[R_1](Rel)$$

$$Rel_2 := (R_2, Q_2), \quad Q_2 := \pi[R_2](Rel)$$

Die Zerlegung von R in R_1 und R_2 ist verlustlos,

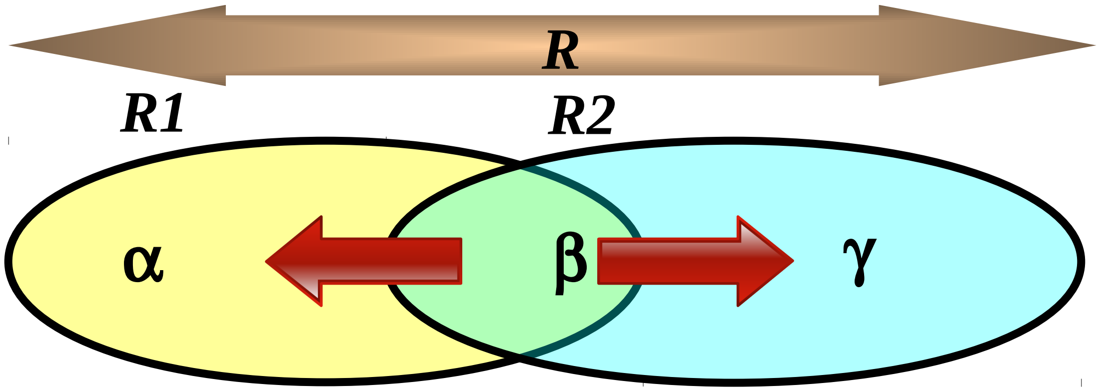
falls für jede mögliche (gültige) Ausprägung Q gilt:

$$R = R_1 \bowtie R_2$$

Hinreichende Bedingung für die Verlustlosigkeit einer Zerlegung

$R_1 \cap R_2 \rightarrow R_1$ oder

$R_1 \cap R_2 \rightarrow R_2$



Biertrinker-Beispiel

Biertrinker		
Lokal	***Gast***	***Bier***
Einstein	Gniffke	Pils
Einstein	Anders	Hefeweizen
Mephisto	Gniffke	Hefeweizen

Biertrinker-Beispiel: Verlustige Zerlegung

Biertrinker		
Lokal	***Gast***	***Bier***
Einstein	Gniffke	Pils
Einstein	Anders	Hefeweizen
Mephisto	Gniffke	Hefeweizen

Besucht := π [*Lokal*, *Gast*] (*Biertrinker*)	
Lokal	***Gast***
Einstein	Gniffke
Einstein	Anders
Mephisto	Gniffke

Trinkt := π [*Gast*, *Bier*] (*Biertrinker*)	
Gast	***Bier***
Gniffke	Pils
Anders	Hefeweizen
Gniffke	Hefeweizen

Besucht ⋈ *Trinkt*		
Lokal	***Gast***	***Bier***
Einstein	Gniffke	Pils
Einstein	**Gniffke**	**Hefeweizen**
Einstein	Anders	Hefeweizen
Mephisto	**Gniffke**	**Pils**
Mephisto	Gniffke	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} $\rightarrow$ {Bier}

Wohingegen keine der zwei möglichen, die Verlustlosigkeit garantierenden FDs gelten

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

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

  - (damit sich die Kellner darauf einstellen können?)

Verlustfreie Zerlegung

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

Väter **:=** π **[*Vater, Kind*] (*Eltern*)**	
Vater	***Kind***
Johann	Else
Johann	Theo
Heinz	Cleo

Mütter **:=** π **[*Mutter, Kind*] (*Eltern*)**	
Mutter	***Kind***
Martha	Else
Maria	Theo
	o

Väter $\bowtie$ ***Mütter***		
Vater	*Mutter*	*Kind*
Johann	Martha	Else
Johann	Maria	Theo
Heinz	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} $\rightarrow$ {Mutter}
- {Kind} $\rightarrow$ {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 $R_1, \ldots, R_n$

$F_R = (F_{R1} \cup \ldots \cup F_{Rn})$ bzw. $F_R^+ = (F_{R1} \cup \ldots \cup F_{Rn})^+$

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} $\rightarrow$ {Ort, BLand}
- {Straße, Ort, BLand} $\rightarrow$ {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

Straßen := π [PLZ, Straße] (PLZverzeichnis)	
PLZ	***Straße***
15234	Goethestraße
60313	Goethestraße
60437	Galgenstraße

Orte := π [Ort, BLand] (PLZverzeichnis)		
Ort	*BLand*	***PLZ***
Frankfurt	Hessen	60313
Frankfurt	Hessen	60437
Frankfurt	Brandenburg	15234

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

Zerlegung der Relation PLZverzeichnis

PLZverzeichnis			
Ort	***BLand***	***Straße***	*PLZ*
Frankfurt	Hessen	Goethestrasse	60313
Frankfurt	Hessen	Galgenstrasse	60437
Frankfurt	Brandenburg	Goethestrasse	15234

***Straßen* := π [*PLZ*, *Straße*] (*PLZverzeichnis*)**	
PLZ	***Straße***
15234	Goethestrasse
60313	Goethestrasse
15235	**Goethestrasse**
60437	Galgenstrasse

***Orte* := π [*Ort*, *BLand*] (*PLZverzeichnis*)**		
Ort	*BLand*	***PLZ***
Frankfurt	Hessen	60313
Frankfurt	Hessen	60437
Frankfurt	Brandenburg	15234
Frankfurt	**Brandenburg**	**15235**

Zerlegung der Relation PLZverzeichnis

***Straßen* := π [*PLZ*, *Straße*] (*PLZverzeichnis*)**	
PLZ	***Straße***
15234	Goethestraße
60313	Goethestraße
15235	**Goethestraße**
60437	Galgenstraße

***Orte* := π [*Ort*, *BLand*] (*PLZverzeichnis*)**		
Ort	*BLand*	***PLZ***
Frankfurt	Hessen	60313
Frankfurt	Hessen	60437
Frankfurt	Brandenburg	15234
Frankfurt	**Brandenburg**	**15235**

Straßen* ⋈ *Orte			
Ort	*BLand*	*Straße*	*PLZ*
Frankfurt	Hessen	Goethestraße	60313
Frankfurt	Hessen	Galgenstraße	60437
Frankfurt	**Brandenburg**	**Goethestraße**	**15235**
Frankfurt	Brandenburg	Goethestraße	15234

Die FD {Straße, Ort, BLand} → {PLZ} wird verletzt

Normalformen: informeller Überblick

1NF: Das Modell ist strikt relational

2NF: Keine Abhängigkeit von Teilschlüssel

3NF: Keine Kettenabhängigkeiten

BCNF: Keine Abhängigkeit von Teilschlüssel, auch nicht in Kandidatenschlüssel

4NF: Beschreibt mehrwertige Abhängigkeiten

5NF: Relation lässt sich nicht weiter aufspalten, ohne dass Information verloren geht

Erste Normalform

Nur atomare Domänen

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

1NF

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

Exkurs: NF²-Relationen

Non-First Normal-Form-Relationen

Geschachtelte Relationen (Oracle: Nested Tables)

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

Zweite Normalform (2NF)

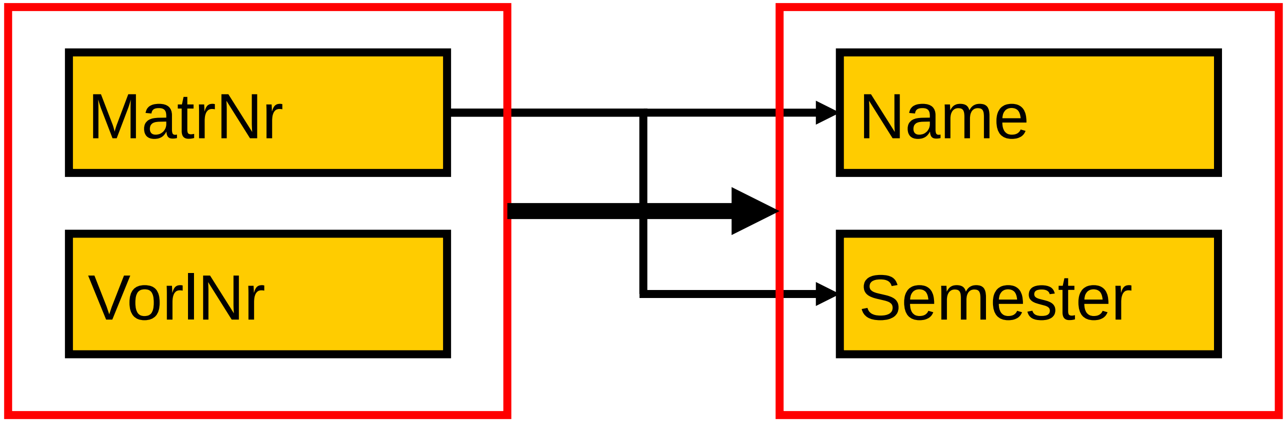
Eine Relation R mit zugehörigen FDs F_R ist in zweiter Normalform, falls jedes Nichtschlüssel-Attribut $A \in R$ voll funktional abhängig ist von jedem Kandidatenschlüssel der Relation.

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 2NF

- {MatrNr} → {Name}
- {MatrNr} → {Semester}

Zweite Normalform: Beispiel



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?

Lösung: Zerlegung in zwei Relationen

- hören: {[MatrNr, VorlNr]}
- Studenten: {[MatrNr, Name, Semester]}

Beide Relationen sind in 2NF

Dritte Normalform (3NF)

Ein Relationenschema R ist in dritter Normalform, wenn für jede für R geltende funktionale Abhängigkeit der Form $\alpha \rightarrow B$ mit $B \in R$ mindestens eine von drei Bedingungen gilt:

- $B \in \alpha$, d.h., die FD ist trivial oder
- Das Attribut B ist ein Schlüsselattribut oder
- α ist Superschlüssel von R

Studenten				
MatrNr	Fachbereich	Standort	Name	Semester
26120	Informatik	Koblenz	Fichte	4
27550	Informatik	Koblenz	Meyer	8
27551	Psychologie	Landau	Carnap	5

Studenten ist **nicht** in 3NF

{Fachbereich} → {Standort}

Zerlegung mit dem Synthesealgorithmus

Ziel:

Zerlegung von R mit vorgegebenen FDs F
in $R_1, ..., R_n$, wobei folgende Kriterien erfüllt sind:

- Zerlegung von R in $R_1, ..., R_n$ erfolgt verlustlos
- Zerlegung $R_1, ..., R_n$ ist abhängigkeitserhaltend
- Alle $R_1, ..., R_n$ sind in 3NF

Synthesealgorithmus

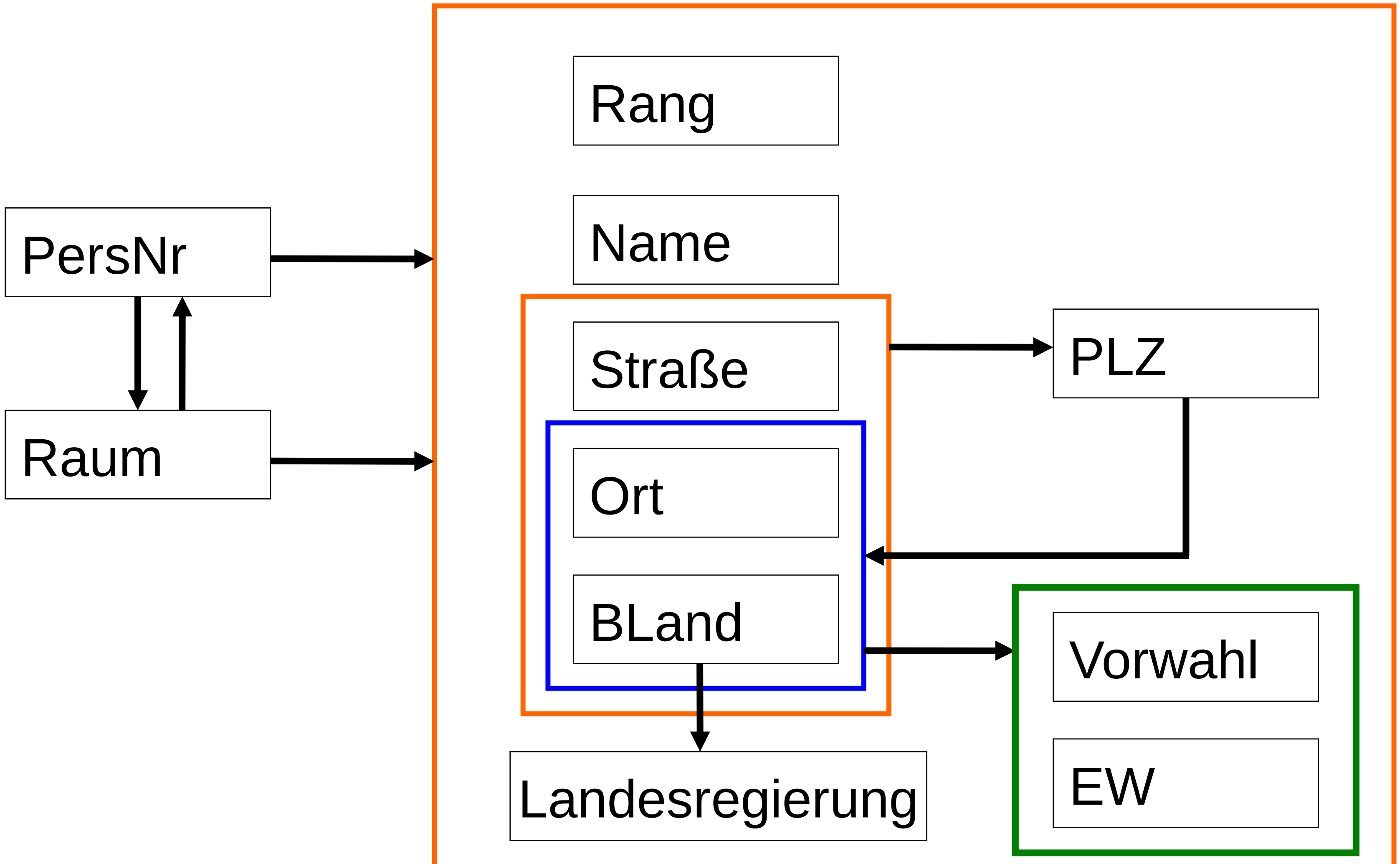
1) Bestimme die kanonische Überdeckung Fc zu F. Wiederholung:
   a. Linksreduktion
   b. Rechtsreduktion
   c. Entfernung von FDs der Form $\alpha \rightarrow \emptyset$
   d. Zusammenfassung gleicher linker Seiten

2) Für jede funktionale Abhängigkeit $\alpha \to \beta \in Fc$:
   - Kreiere ein Relationenschema $R\alpha := \alpha \cup \beta$
   - Ordne R_α die FDs $F_\alpha := \{\alpha' \to \beta' \in Fc \mid \alpha' \cup \beta' \subseteq R_\alpha\}$ zu.

3) Falls eines der in Schritt 2. erzeugten Schemata einen Kandidatenschlüssel von R bzgl. Fc enthält, sind wir fertig. Sonst wähle einen Kandidatenschlüssel $k \subseteq R$ aus und definiere folgendes Schema:
   a. $R_k := k$
   b. $F_k := \emptyset$

4) Eliminiere diejenigen Schemata R_α, die in einem anderen Relationenschema $R\alpha'$ enthalten sind, d.h.,
   - $R_\alpha \subseteq R_{\alpha'}$

Anwendung des Synthesealgorithmus



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]}

2. und 6. fallen weg (Teilrelationen von 1. und 3.)

Gesamtkandidatenschlüssel {PersNr} is bereits Schlüssel einer Relation

Boyce-Codd-Normalform (BCNF)

BCNF = eine weitere Verschärfung der 3NF

Ein Relationenschema R mit FDs F ist in BCNF, wenn für jede für R geltende funktionale Abhängigkeit der Form $\alpha \rightarrow \beta \in F$ gilt:

- entweder $\beta \subseteq \alpha$, d.h., die Abhängigkeit ist trivial,
- oder α ist ein Superschlüssel von R

Man kann jede Relation verlustlos in BCNF-Relationen zerlegen

Manchmal lässt sich dabei die Abhängigkeiterhaltung aber nicht erzielen

Beispiel: Relation in 3NF, aber nicht in BCNF

Städte: {[Ort, BLand, Ministerpräsident, EW]}

Geltende FDs:

- {Ort, BLand} → {EW}
- {BLand} → {Ministerpräsident}
- {Ministerpräsident} → {BLand}

Kandidatenschlüssel:

- {Ort, BLand}
- {Ort, Ministerpräsident}

Dekomposition

Man kann grundsätzlich jedes Relationenschema R mit funktionalen Anhängigkeiten F so in $R_1, \ldots, R_n$ zerlegen, dass gilt:

- $R_1, \ldots, R_n$ ist eine verlustlose Zerlegung von R
- Alle $R_1, \ldots, R_n$ sind in BCNF
- Es kann leider nicht immer erreicht werden, dass die Zerlegung $R_1, \ldots, R_n$ abhängigkeitserhaltend 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 $(\alpha \to \beta)$ mit
  - $\alpha \cap \beta = \emptyset$
  - $\neg(\alpha \to R_i)$
- 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
  - $Z := (Z - \{R_i\}) \cup \{R_{i1}\} \cup \{R_{i2}\}$

Kriterien für die Verlustlosigkeit einer Zerlegung

$$Rel := (R, Q) \qquad R = R_1 \cup R_2$$

$$Rel_1 := (R_1, Q_1), \quad Q_1 := \pi[R_1](Rel)$$

$$Rel_2 := (R_2, Q_2), \quad Q_2 := \pi[R_2](Rel)$$

Die Zerlegung von R in R_1 und R_2 ist verlustlos,

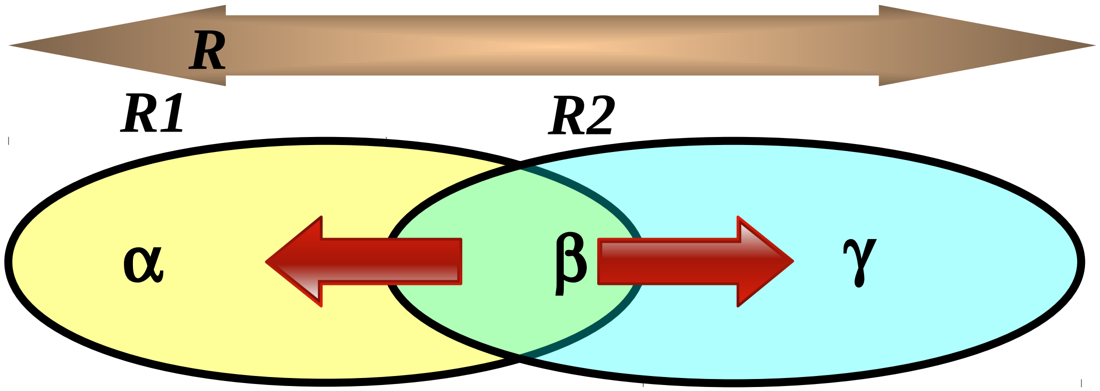
falls für jede mögliche (gültige) Ausprägung Q gilt:

$$R = R_1 \bowtie R_2$$

Hinreichende Bedingung für die Verlustlosigkeit einer Zerlegung

$R_1 \cap R_2 \rightarrow R_1$ oder

$R_1 \cap R_2 \rightarrow R_2$



Dekomposition der Relation Städte in BCNF-Relationen

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

Geltende FDs:

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

R_{i1}:

- Regierungen: {[BLand, Ministerpräsident/in]}

R_{i2}:

- Städte: {[Ort, BLand, EW]}

Zerlegung ist verlustlos und auch abhängigkeitserhaltend

Dekomposition des PLZverzeichnis in BCNF-Relationen

PLZverzeichnis: {[Straße, Ort, BLand, PLZ]}

Funktionale Abhängigkeiten:

- {PLZ} → {Ort, BLand}
- {Straße, Ort, BLand} → {PLZ}

Betrachte die Zerlegung

- Orte: {[PLZ, Ort, BLand]}
- Straßen: {[PLZ, Straße]}

Diese Zerlegung

- ist verlustlos aber
- Nicht abhängigkeitserhaltend

Zerlegung der Relation PLZverzeichnis

PLZverzeichnis			
Ort	***BLand***	***Straße***	*PLZ*
Frankfurt	Hessen	Goethestrasse	60313
Frankfurt	Hessen	Galgenstrasse	60437
Frankfurt	Brandenburg	Goethestrasse	15234

***Straßen* := π [*PLZ*, *Straße*] (*PLZverzeichnis*)**	
PLZ	***Straße***
15234	Goethestrasse
60313	Goethestrasse
15235	**Goethestrasse**
60437	Galgenstrasse

***Orte* := π [*Ort*, *BLand*] (*PLZverzeichnis*)**		
Ort	*BLand*	***PLZ***
Frankfurt	Hessen	60313
Frankfurt	Hessen	60437
Frankfurt	Brandenburg	15234
Frankfurt	**Brandenburg**	**15235**

Zerlegung der Relation PLZverzeichnis

***Straßen* := π [*PLZ, Straße*] (*PLZverzeichnis*)**	
PLZ	***Straße***
15234	Goethestrasse
60313	Goethestrasse
15235	**Goethestrasse**
60437	Galgenstrasse

***Orte* := π [*Ort, BLand*] (*PLZverzeichnis*)**		
Ort	*BLand*	***PLZ***
Frankfurt	Hessen	60313
Frankfurt	Hessen	60437
Frankfurt	Brandenburg	15234
Frankfurt	**Brandenburg**	**15235**

Strassen* ⋈ *Orte			
Ort	*BLand*	*Straße*	*PLZ*
Frankfurt	Hessen	Goethestrasse	60313
Frankfurt	Hessen	Galgenstrasse	60437
Frankfurt	**Brandenburg**	**Goethestrasse**	**15235**
Frankfurt	Brandenburg	Goethestrasse	15234

Die FD {Straße, Ort, BLand} $\rightarrow$ {PLZ} wird verletzt

Fragen ?

kunegis@uni-koblenz.de

http://west.uni-koblenz.de/teaching/ws1213/datenbanken