Kapitel 6
Relationale Entwurfstheorie

Funktionale Abhängigkeiten
Normalformen
Normalisierung durch Dekomposition

Ziele der relationalen Entwurfstheorie

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

Funktionale Abhängigkeiten

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

- Seien $\alpha \subseteq \mathcal{R}$, $\beta \subseteq \mathcal{R}$
- $\alpha \to \beta$ 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

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

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

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

Schlüssel

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

Schlüsselbestimmung

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

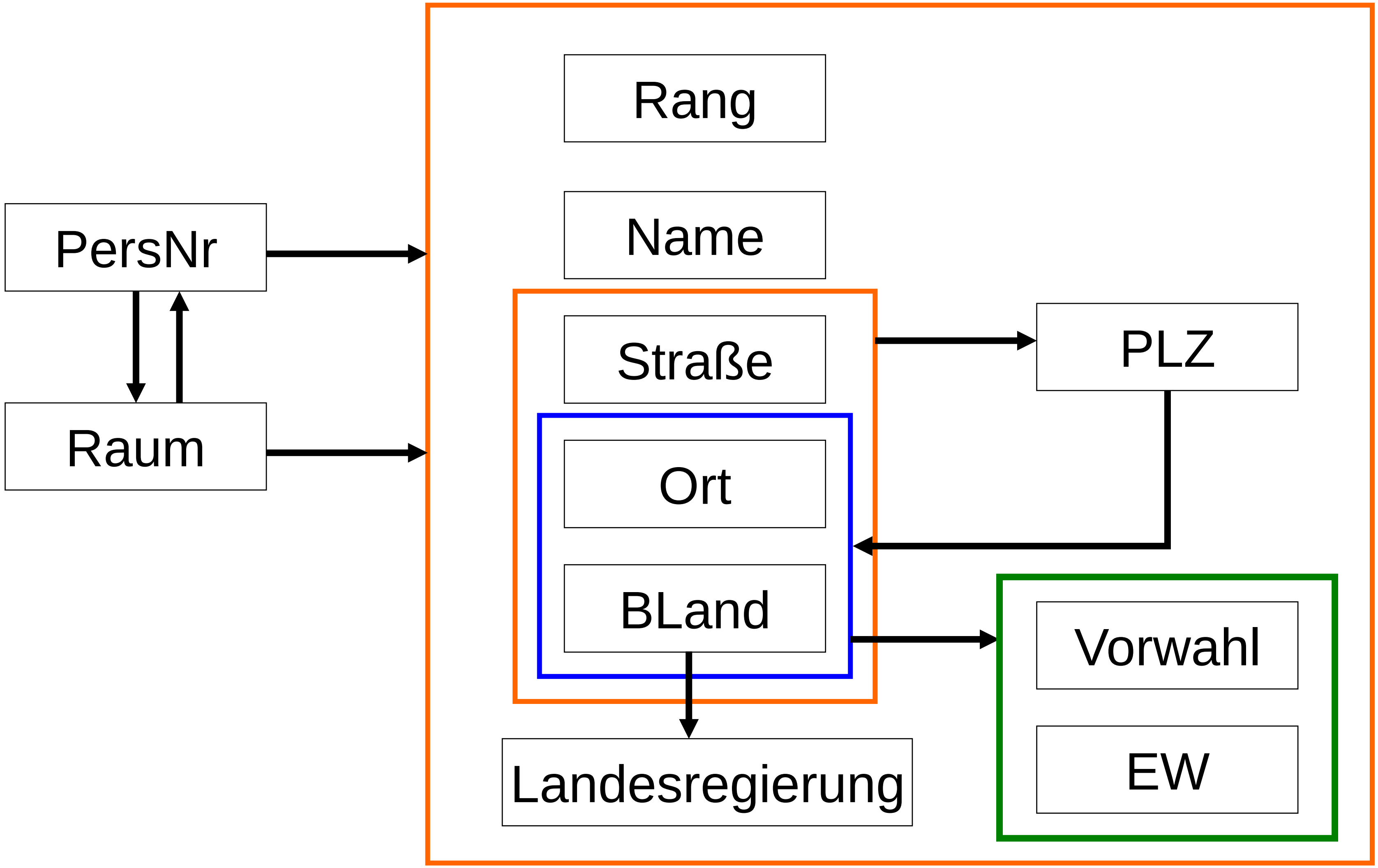
- Kandidaten-schlüssel von *Städte*:
  - {Name,BLand}
  - {Name,Vorwahl}

- Beachte, dass 2 kleinere Städte dieselbe Vorwahl haben können

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



„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 $R1$, ..., Rn der neuen Relationenschemata $\mathcal{R}1$, .., $\mathcal{R}n$ rekonstruierbar sein.

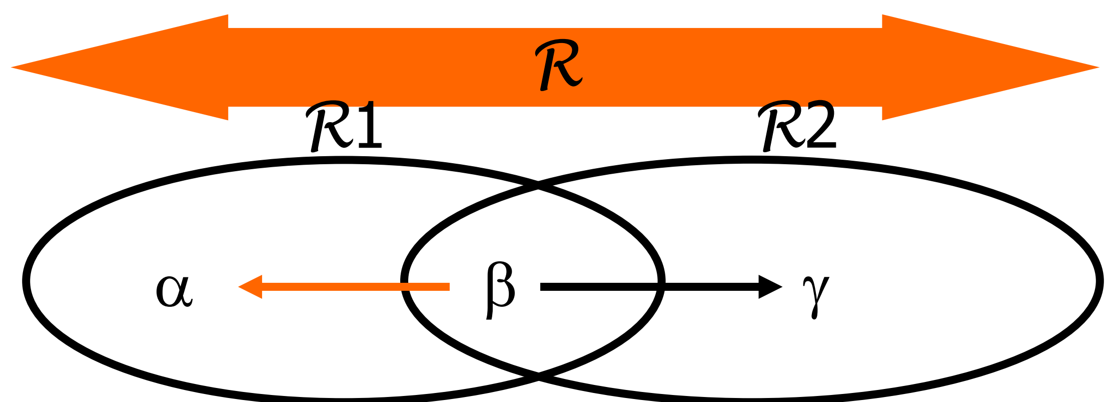
2. Abhängigkeitserhaltung
   - Die für $\mathcal{R}$ geltenden funktionalen Anhängigkeiten müssen auf die Schemata $\mathcal{R}1$, ..., $\mathcal{R}n$ übertragbar sein.

Kriterien für die Verlustlosigkeit einer Zerlegung

- $\mathcal{R} = \mathcal{R}1 \cup \mathcal{R}2$
    - $R1 := \Pi_{\mathcal{R}1}(R)$
    - $R2 := \Pi_{\mathcal{R}2}(R)$

- 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 = R1 A R2

- Hinreichende Bedingung für die Verlustlosigkeit einer Zerlegung
    - $(\mathcal{R}1 \cap \mathcal{R}2) \rightarrow \mathcal{R}1$ oder
    - $(\mathcal{R}1 \cap \mathcal{R}2) \rightarrow \mathcal{R}2$



Biertrinker-Beispiel

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

„Verlustige“ Zerlegung

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

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

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

Besucht	
Kneipe	*Gast*
Kowalski	Kemper
Kowalski	Eickler
Innsteg	Kemper

Trinkt	
Gast	*Bier*
Kemper	Pils
Eickler	Hefeweizen
Kemper	Hefeweizen

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

$\Pi_{....}$

Besucht	
Kneipe	*Gast*
Kowalski	Kemper
Kowalski	Eickler
Innsteg	Kemper

Trinkt	
Gast	*Bier*
Kemper	Pils
Eickler	Hefeweizen
Kemper	Hefeweizen

A

$\neq$

Besucht A Trinkt		
Kneipe	*Gast*	*Bier*
Kowalski	Kemper	Pils
Kowalski	Kemper	Hefeweizen
Kowalski	Eickler	Hefeweizen
Innsteg	Kemper	Pils
Innsteg	Kemper	Hefeweizen

Erläuterung des Biertrinker-Beispiels

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

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

- Das liegt daran, dass die Leute (insbes. Kemper) in unterschiedlichen Kneipen unterschiedliches Bier trinken. In derselben Kneipe aber immer das gleiche Bier
  - (damit sich die KellnerInnen darauf einstellen können?)

Verlustfreie Zerlegung

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

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

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

Väter	
Vater	*Kind*
Johann	Else
Johann	Theo
Heinz	Cleo

Mütter	
Mutter	*Kind*
Martha	Else
Maria	Theo
Martha	Cleo

Erläuterung der verlustfreien Zerlegung der Eltern-Relation

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

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

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

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

Boyce-Codd-Normalform

- Die Boyce-Codd-Normalform (BCNF) ist nochmals eine Verschärfung der 3 NF.
- Ein Relationenschema $\mathcal{R}$ mit FDs F ist in BCNF, wenn für jede für $\mathcal{R}$ geltende funktionale Abhängigkeit der Form $\alpha \rightarrow \beta \in F$ und mindestens eine von zwei Bedingungen gilt:
  - $\beta \subseteq \alpha$, d.h., die Abhängigkeit ist trivial oder
  - α ist Superschlüssel von $\mathcal{R}$

- Man kann jede Relation verlustlos in BCNF-Relationen zerlegen
- Manchmal lässt sich dabei die Abhängigkeiterhaltung aber nicht erzielen

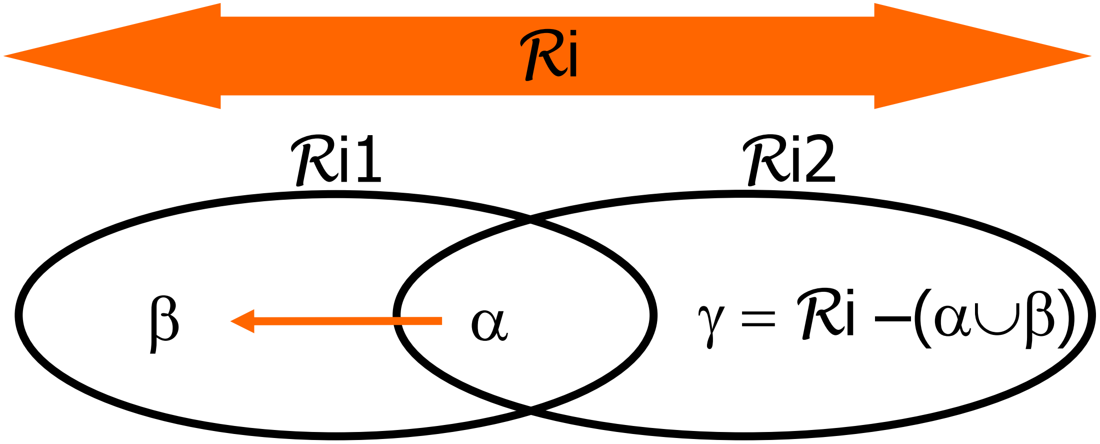
Dekomposition

- Man kann grundsätzlich jedes Relationenschema $\mathcal{R}$ mit funktionalen Anhängigkeiten F so in $\mathcal{R}1, ..., \mathcal{R}n$ zerlegen, dass gilt:
  - $\mathcal{R}1, ..., \mathcal{R}n$ ist eine verlustlose Zerlegung von $\mathcal{R}$.
  - Alle $\mathcal{R}1, ..., \mathcal{R}n$ sind in BCNF.
  - Es kann leider nicht immer erreicht werden, dass die Zerlegung $\mathcal{R}1, ..., \mathcal{R}n$ abhängigkeitserhaltend ist.

Dekompositions-Algorithmus

- Starte mit $Z = \{\mathcal{R}\}$
- Solange es noch ein Relationenschema $\mathcal{R}i$ in Z gibt, das nicht in BCNF ist, mache folgendes:
  - Es gibt also eine für $\mathcal{R}i$ geltende nicht-triviale funktionale Abhängigkeit $(\alpha \to \beta)$ mit
    - $\alpha \cap \beta = \varnothing$
    - $\neg(\alpha \to \mathcal{R}i)$
  - 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 $\mathcal{R}i1 := \alpha \cup \beta$ und $\mathcal{R}i2 := \mathcal{R}i - \beta$
  - Entferne $\mathcal{R}i$ aus Z und füge $\mathcal{R}i1$ und $\mathcal{R}i2$ ein, also
    - $Z := (Z - \{\mathcal{R}i\}) \cup \{\mathcal{R}i1\} \cup \{\mathcal{R}i2\}$

Veranschaulichung der Dekomposition



Dekomposition der Relation Städte in BCNF-Relationen

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

- $\mathcal{R}$i1:
  - Regierungen: {[BLand, Ministerpräsident/in]}

- $\mathcal{R}$i2:
  - Städte: {[Ort, BLand, EW]}

- Zerlegung ist verlustlos und auch abhängigkeitserhaltend