

Datenbanksysteme 2012

Kapitel 12: Relationale Entwurfstheorie

Oliver Vornberger

Institut für Informatik
Universität Osnabrück

Funktionale Abhängigkeiten

$$\alpha \rightarrow \beta$$

β ist funktional abhängig von α

$$\forall r, t \in R : r.\alpha = t.\alpha \Rightarrow r.\beta = t.\beta$$

R			
A	B	C	D
a_4	b_2	c_4	d_3
a_1	b_1	c_1	d_1
a_1	b_1	c_1	d_2
a_2	b_2	c_3	d_2
a_3	b_2	c_4	d_3

Es gilt:

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

$$\{A\} \rightarrow \{C\}$$

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

Es gilt nicht:

$$\{B\} \rightarrow \{C\}.$$

Test auf Abhängigkeit

- 1.) Sortiere R nach α -Werten
- 2.) Teste, ob bei gleichen α
auch gleiche β vorliegen

Schreibweise

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

schreiben wir $CD \rightarrow B$

statt $\alpha \cup \beta$

schreiben wir $\alpha\beta$

Schlüssel

In dem Relationenschema \mathcal{R}

ist $\alpha \subseteq \mathcal{R}$ ein *Superschlüssel* falls gilt $\alpha \rightarrow \mathcal{R}$

β ist voll funktional abhängig von α , falls gilt

$$\alpha \rightarrow \beta$$

$$\forall A \in \alpha : \alpha - \{A\} \not\rightarrow \beta$$

$\alpha \subseteq \mathcal{R}$ heisst *Schlüsselkandidat* falls gilt

\mathcal{R} ist voll funktional abhängig von α

Primärschlüssel = einer der Schlüsselkandidaten

Relation Städte

Ort	Bland	Vorwahl	EW
Frankfurt	Hessen	069	650000
Frankfurt	Brandenburg	0335	84000
München	Bayern	089	1200000
Passau	Bayern	0851	50000
...

Vorwahl wechselt nicht im Ort, mehrere Orte können dieselbe Vorwahl haben
Orte erstrecken sich nicht über Ländergrenzen

Schlüsselkandidaten:

{Ort, BLand}

{Ort, Vorwahl}

Relation ProfessorenAdr

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

Abhängigkeiten:

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

{Ort, BLand} → {Vorwahl}

{PLZ} → {BLand, Ort}

{Ort, BLand, Straße} → {PLZ}

{BLand} → {Landesregierung}

{Raum} → {PersNr}

davon abgeleitet:

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

{PLZ} → {Landesregierung}

Hülle von F

Gegeben:

Menge von funktionalen Abhängigkeiten F

Gesucht:

$F^+ :=$ Menge der aus F ableitbaren Abhängigkeiten

Armstrong Axiome

- Reflexivität: Aus $\beta \subseteq \alpha$ folgt: $\alpha \rightarrow \beta$
- Verstärkung: Aus $\alpha \rightarrow \beta$ folgt: $\alpha\gamma \rightarrow \beta\gamma$ für $\gamma \subseteq U$
- Transitivität: Aus $\alpha \rightarrow \beta$
und $\beta \rightarrow \gamma$ folgt: $\alpha \rightarrow \gamma$

Die Armstrong-Axiome sind

- *sound* (korrekt)
- *complete* (vollständig)

Weitere Axiome

- Vereinigung: Aus $\alpha \rightarrow \beta$
und $\alpha \rightarrow \gamma$
folgt: $\alpha \rightarrow \beta\gamma$
- Dekomposition: Aus $\alpha \rightarrow \beta\gamma$
folgt: $\alpha \rightarrow \beta$
und $\alpha \rightarrow \gamma$
- Pseudotransitivität: Aus $\alpha \rightarrow \beta$
und $\gamma\beta \rightarrow \delta$
folgt: $\alpha\gamma \rightarrow \delta$

Beispiel für Armstrong-Axiome

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

abzuleiten: {PLZ} → {Landesregierung}

{PLZ} → {BLand}	(Dekomposition)
{BLand} → {Landesregierung}	(als FD gegeben)
{PLZ} → {Landesregierung}	(Transitivität)

Wunsch: alle Abhängigkeiten erfahren

Gegeben:

Menge von funktionalen Abhängigkeiten F

Gesucht:

$F^+ :=$ Menge der aus F ableitbaren Abhängigkeiten

Aber: F^+ kann auch bei kleinem F recht groß werden

Abschluss einer Attribut-Menge

$$\alpha^+ := \{ \beta \subseteq U \mid \alpha \rightarrow \beta \in F^+ \}$$

Satz:

$\alpha \rightarrow \beta$ folgt aus Armstrongaxiomen $\Leftrightarrow \beta \in \alpha^+$.

Algorithmus zur Bestimmung von α^+ :

$$X^0 := \alpha$$

$$X^{i+1} := X^i \cup \gamma \text{ falls } \beta \rightarrow \gamma \in F \wedge \beta \subseteq X^i$$

Abbruch, falls unverändert

Beispiel für Abschluss einer Attribut-Menge

Sei $U = \{A, B, C, D, E, G\}$

Sei $F = \{AB \rightarrow C, C \rightarrow A, BC \rightarrow D, ACD \rightarrow B, D \rightarrow EG, BE \rightarrow C, CG \rightarrow BD, CE \rightarrow AG\}$

Sei $\alpha = \{B, D\}$

$X^0 = BD$

$X^1 = BDEG$

$X^2 = BCDEG$

$X^3 = ABCDEG$

$X^4 = ABCDEG$

\Rightarrow Abbruch

Also: $\alpha^+ = ABCDEG$

Äquivalenz von funktionalen Abhängigkeiten

$$F \equiv G \Leftrightarrow F^+ = G^+$$

Algorithmus:

Teste für jede Abhängigkeit $\alpha \rightarrow \beta \in F$, ob gilt:

$$\alpha \rightarrow \beta \in G^+, \text{ d. h. } \beta \subseteq \alpha^+ \text{ bzgl. } G$$

Teste für jede Abhängigkeit $\gamma \rightarrow \delta \in G$, ob gilt:

$$\gamma \rightarrow \delta \in F^+, \text{ d. h. } \delta \subseteq \gamma^+ \text{ bzgl. } F$$

Minimale Menge von funktionalen Abhängigkeiten

- Jede rechte Seite hat nur ein Attribut.
- Weglassen einer Abhängigkeit aus F verändert F^+ .
- Weglassen eines Attributs in der linken Seite verändert F^+ .

Algorithmus:

- Aufsplitten der rechten Seiten.
- Probeweises Entfernen von Regeln bzw. von Attributen auf der linken Seite.

Beispiel für Äquivalenz

$$U = \{ A, B, C, D, E, G \}$$

Aufspalten der rechten Seiten:

$$F = \left\{ \begin{array}{l} AB \rightarrow C, \\ C \rightarrow A, \\ BC \rightarrow D, \\ ACD \rightarrow B, \\ D \rightarrow EG, \\ BE \rightarrow C, \\ CG \rightarrow BD, \\ CE \rightarrow AG \end{array} \right\}$$

$$\begin{array}{l} AB \rightarrow C \\ C \rightarrow A \\ BC \rightarrow D \\ ACD \rightarrow B \\ D \rightarrow E \\ D \rightarrow G \\ BE \rightarrow C \\ CG \rightarrow B \\ CG \rightarrow D \\ CE \rightarrow A \\ CE \rightarrow G \end{array}$$

Entfernen von Redundanz

$AB \rightarrow C$

$C \rightarrow A$

$BC \rightarrow D$

$ACD \rightarrow B$

$D \rightarrow E$

$D \rightarrow G$

$BE \rightarrow C$

$CG \rightarrow B$

$CG \rightarrow D$

$CE \rightarrow A$

$CE \rightarrow G$

$CE \rightarrow A$ ist redundant wegen

$C \rightarrow A$

$CG \rightarrow B$ ist redundant wegen

$CG \rightarrow D$

$C \rightarrow A$

$ACD \rightarrow B$

$ACD \rightarrow B$ kann gekürzt werden zu

$CD \rightarrow B$ wegen

$C \rightarrow A$

Schlechte Relationenschemata

PersNr	Name	Rang	Raum	VorlNr	Titel	SWS
2125	Sokrates	C4	226	5041	Ethik	4
2125	Sokrates	C4	226	5049	Mäutik	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-Anomalie Angaben zu Professor mehrfach gespeichert
- Insert Anomalie Professor nur mit Vorlesung einfügen
- Delete-Anomalie Entfernen von Vorlesung entfernt Professor

Normalisierung

Zerlegung eines Schemas \mathcal{R} in Schemata $\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_n$ mit

- Verlustlosigkeit:

Die in der ursprünglichen Ausprägung R des Schemas \mathcal{R} enthaltenen Informationen

müssen aus den Ausprägungen R_1, \dots, R_n

der neuen Schemata $\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_n$ rekonstruierbar sein

- Abhängigkeitserhaltung:

Die für \mathcal{R} geltenden funktionalen Abhängigkeiten

müssen auf die $\mathcal{R}_1, \dots, \mathcal{R}_n$ übertragbar sein.

Zerlegung in zwei Relationenschemata

$$\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$$

$$R_1 := \Pi_{\mathcal{R}_1} (R)$$

$$R_2 := \Pi_{\mathcal{R}_2} (R)$$

Eine Zerlegung von \mathcal{R} in \mathcal{R}_1 und \mathcal{R}_2 heißt *verlustlos*, falls für jede gültige Ausprägung R von \mathcal{R} gilt:

$$R = R_1 \bowtie R_2$$

Relation Biertrinker

Biertrinker

Kneipe	Gast	Bier
Stiefel	Wacker	Pils
Stiefel	Sorglos	Hefeweizen
Zwiebel	Wacker	Hefeweizen

Besucht

Kneipe	Gast
Stiefel	Wacker
Stiefel	Sorglos
Zwiebel	Wacker

Gast	Bier
Wacker	Pils
Sorglos	Hefeweizen
Wacker	Hefeweizen

Trinkt

Besucht $\triangleright \triangleleft$ Trinkt

Kneipe	Gast	Pils
Stiefel	Wacker	Pils
Stiefel	Wacker	Hefeweizen
Stiefel	Sorglos	Hefeweizen
Zwiebel	Wacker	Pils
Zwiebel	Wacker	Hefeweizen

\Rightarrow Nicht
verlustlos !

Abhängigkeitsbewahrend

Zerlegung von \mathcal{R} in $\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_n$ heißt abhängigkeitsbewahrend (hüllentreu) falls gilt

$$F_{\mathcal{R}} \equiv (F_{\mathcal{R}_1} \cup \dots \cup F_{\mathcal{R}_n}) \text{ bzw.}$$

$$F_{\mathcal{R}}^+ = (F_{\mathcal{R}_1} \cup \dots \cup F_{\mathcal{R}_n})^+$$

Relation PLZverzeichnis

(PLZ wechselt nicht in Straße)

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

PLZverzeichnis

$\{PLZ\} \rightarrow \{Ort, BLand\}$

$\{Ort, BLand, Straße\}$

$\rightarrow \{PLZ\}$

Orte

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

PLZ	Straße
60313	Goethestraße
60437	Galgenstraße
15234	Goethestraße

Straßen

Frankfurt	Brandenburg	15235
-----------	-------------	-------

15235	Goethestraße
-------	--------------

verlustlos, da PLZ einziges gemeinsames Attribut und $\{PLZ\} \rightarrow \{Ort, BLand\}$

nicht abhängigkeiterhaltend: wg. $\{Ort, BLand, Straße, \} \rightarrow \{PLZ\}$

Einfügen ok. Aber: Nach Join Problem wg. $\{Ort, BLand, Straße, \} \rightarrow \{PLZ\}$

Erste Normalform (1NF)

Verboten sind mengenwertige Attribute:

Vater	Mutter	Kinder
Johann	Martha	{Else, Lucia}
Johann	Maria	{Theo, Josef}
Heinz	Martha	{Cleo}

Verlangt werden atomare Attribute:

Vater	Mutter	Kind
Johann	Martha	Else
Johann	Martha	Lucia
Johann	Maria	Theo
Johann	Maria	Josef
Heinz	Martha	Cleo

Zweite Normalform (2NF)

Ein Attribut heißt *Primärattribut*, wenn es in mindestens einem Schlüsselkandidaten vorkommt, andernfalls heißt es Nichtprimärattribut.

Ein Relationenschema \mathcal{R} ist in zweiter Normalform falls gilt:

- \mathcal{R} ist in der ersten Normalform
- Jedes Nichtprimär-Attribut $A \in \mathcal{R}$ ist voll funktional abhängig von jedem Schlüsselkandidaten.

Relation Studentenbelegung

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
...

Schlüsselkandidaten:

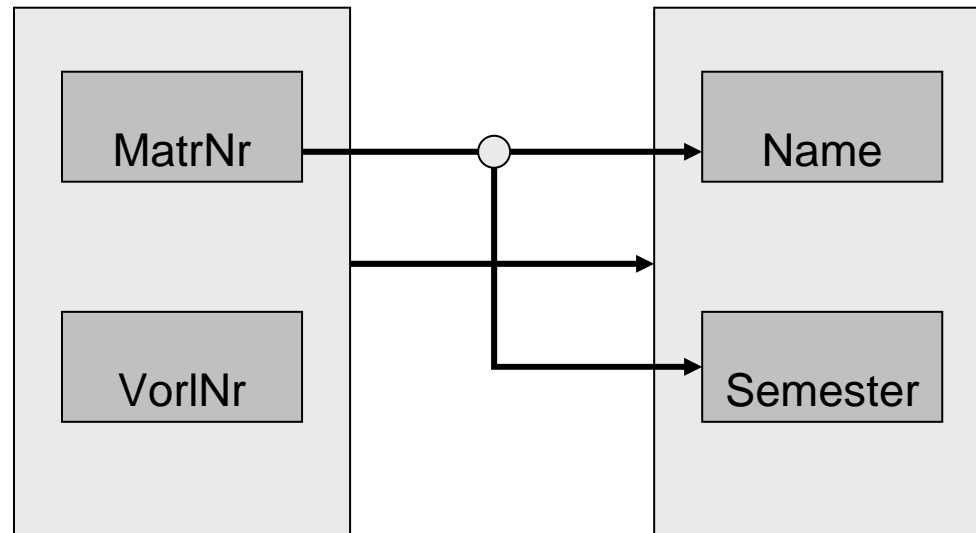
{MatrNr, VorlNr}

Nichtprimärattribute:

{Name, Semester}

Name ist nicht
voll funktional abhängig
von {MatrNr, VorlNr}

⇒ keine 2. Normalform



Relation Hörsaal

Hörsaal

Vorlesung	Dozent	Termin	Raum
Backen ohne Fett	Kant	Mo, 10:15	32/102
Selber Atmen	Sokrates	Mo, 14:15	31/449
Selber Atmen	Sokrates	Di, 14:15	31/449
Schneller Beten	Sokrates	Fr, 10:15	31/449

Schlüsselkandidaten:

{Vorlesung, Termin}

{Dozent, Termin}

{Raum, Termin}

Es gibt keine
Nicht-Primärattribute

⇒ 2. Normalform

Relation Student

Student

MatrNr	Name	Fachbereich	Dekan
29555	Feuerbach	6	Matthies
27550	Schopenhauer	6	Matthies
26120	Fichte	4	Kapphan
25403	Jonas	6	Matthies
28106	Carnap	7	Weingarten

- *Student* in zweiter Normalform
(denn alle Nichtprimärattribute sind voll funktional abhängig von MatrNr)

aber

- Abhängigkeiten zwischen den Nichtprimärattributen,
z. B. hängt *Dekan* von *Fachbereich* ab.

Transitive Abhängigkeit

Gegeben Attributmenge U mit Teilmengen X, Y, Z
 Z heißt transitiv abhängig von X , falls gilt

$$X \cap Z = \emptyset$$

$$\exists Y \subset U : X \cap Y = \emptyset, Y \cap Z = \emptyset$$

$$X \rightarrow Y \rightarrow Z, Y \not\rightarrow X$$

Beispiel:

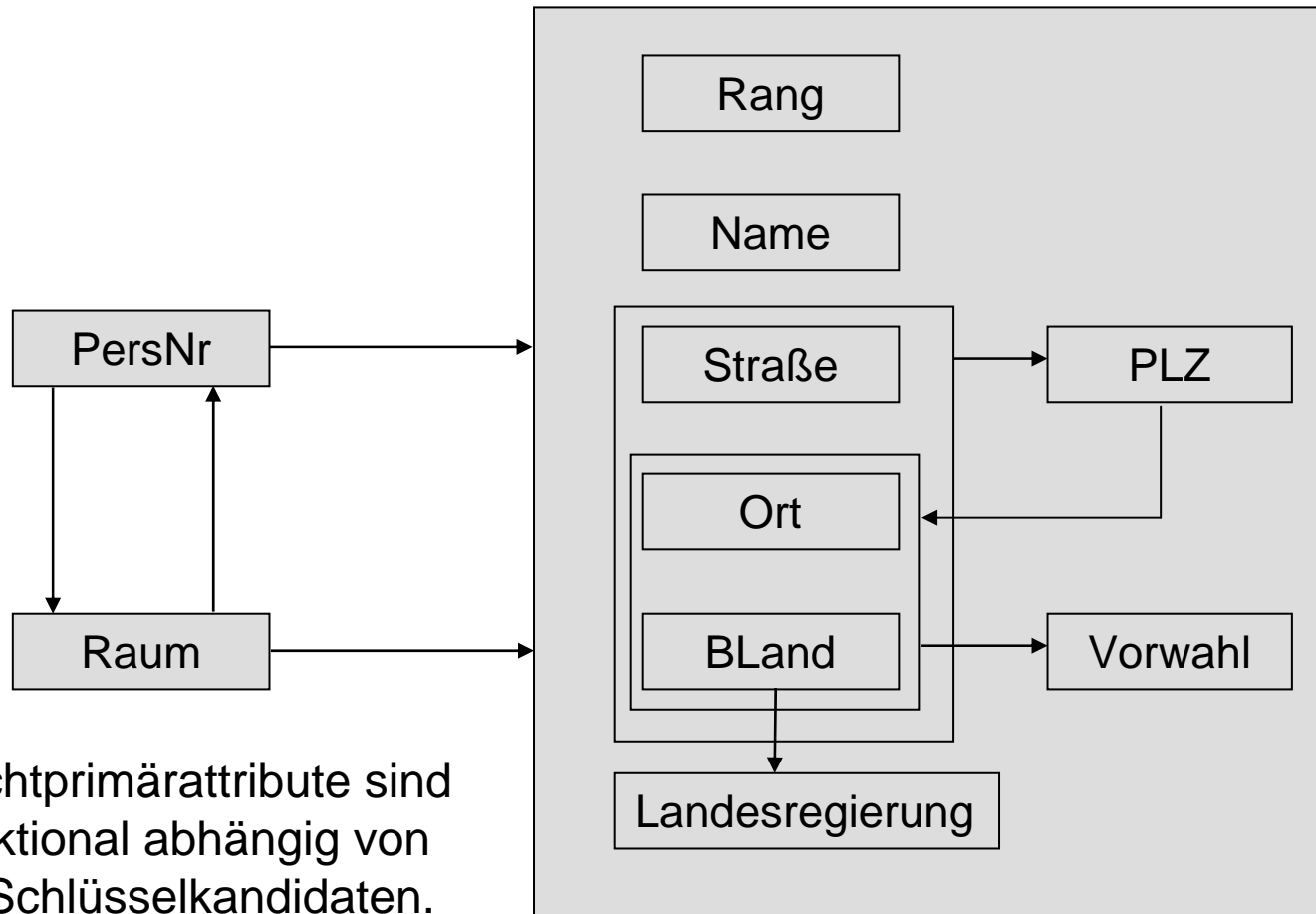
$MatrNr \xrightarrow{\leftarrow} Fachbereich \rightarrow Dekan$

Dritte Normalform (3NF)

\mathcal{R} ist in dritter Normalform

- \mathcal{R} ist in zweiter Normalform
- Jedes Nichtprimärattribut ist nicht-transitiv abhängig von jedem Schlüsselkandidaten

Relation ProfessorenAdr



Alle Nichtprimärattribute sind voll funktional abhängig von jedem Schlüsselkandidaten.

⇒ 2. Normalform

$PersNr \xrightarrow{\neq} \{Ort, BLand\} \rightarrow Vorwahl$ ⇒ nicht in 3. Normalform

Boyce Codd Normalform

\mathcal{R} ist in Boyce Codd Normalform (BCNF):

Für jede funktionale Abhängigkeit $\alpha \rightarrow \beta$ gilt

- $\beta \subseteq \alpha$ (d.h. trivial) oder
- α ist Superschlüssel von \mathcal{R}

Relation Städte

Einziges
Nichtprimär
-attribut

Städte

Ort	BLand	Ministerpräsident	EW
Frankfurt	Hessen	Bouffier	660.000
Frankfurt	Brandenburg	Platzek	70.000
Bonn	NRW	Kraft	300.000
Lotte	NRW	Kraft	14.000

Abhängigkeiten:

$\{\text{Ort, BLand}\} \rightarrow \{\text{EW}\}$

$\{\text{BLand}\} \rightarrow \{\text{Ministerpräsident}\}$

$\{\text{Ministerpräsident}\} \rightarrow \{\text{BLand}\}$

Schlüsselkandidaten:

$\{\text{Ort, BLand}\}$

$\{\text{Ort, Ministerpräsident}\}$

EW ist nicht-transitiv abhängig von Schlüsselkandidaten

\Rightarrow 3. Normalform

$\{\text{Ministerpräsident}\} \rightarrow \{\text{BLand}\}$ ist eine Abhängigkeit, aber

$\{\text{Ministerpräsident}\}$ kein Superschlüssel \Rightarrow nicht BCNF