

6. NORMALISIERUNG VON RELATIONEN

- Einführung
- funktionale Abhängigkeiten
 - Bestimmung von Schlüsselkandidaten
 - Äquivalenzbeziehungen
- Zerlegung von Relationen zur Beseitigung von Anomalien
 - Korrektheitskriterien
- Normalisierung
 - 1NF, atomare Attributwerte
 - 2NF, partielle Abhängigkeiten
 - 3NF, transitive Abhängigkeiten
 - BCNF (Boyce-Codd-Normalform), Determinanten
- Probleme der Normalisierung



EINFÜHRUNG

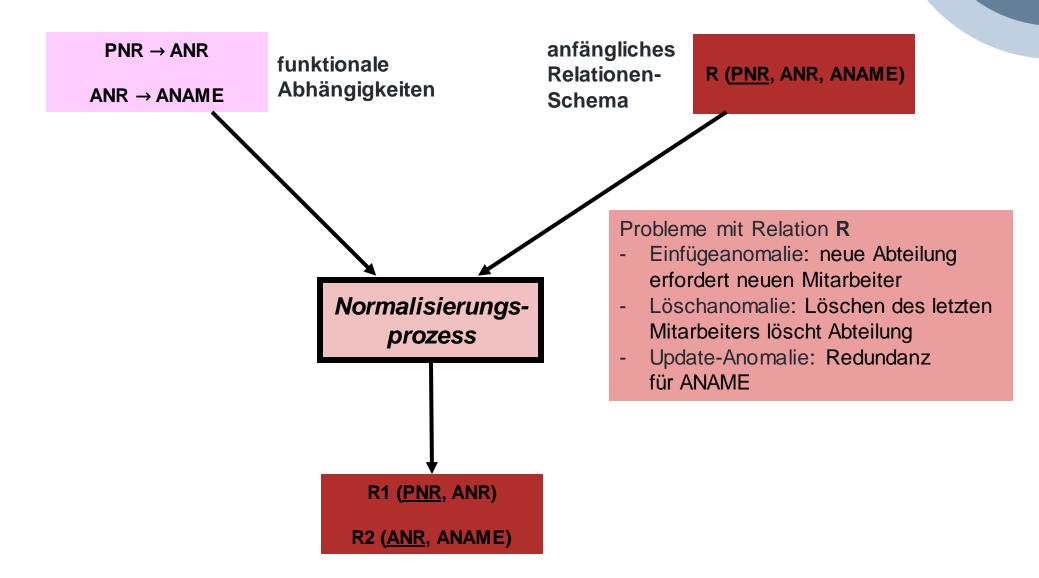
- Ziel: Theoretische Grundlage für "gute" relationale DB-Schemas
- Normalisierung von Relationen: Verbesserung eines gegebenen Schema-Entwurfs
 - teilweise Formalisierung von "Güte" eines Schemas
 - semiformales Verfahren zur Korrektur schlechter Schemas
- Merkmale eines schlechten DB-Schema-Entwurfs
 - implizite Darstellung von Informationen
 - Redundanzen
 - potenzielle Inkonsistenz (Änderungsanomalien)
 - Einfügeanomalien
 - Löschanomalien ...

oft hervorgerufen durch

- "Vermischung" von Entities,
- Zerlegung und wiederholte Speicherung von Entities, ...



NORMALISIERUNG VON RELATIONEN (BSP.)



normalisierte Relationen-Schemata



DEFINITIONEN UND BEGRIFFE

Konventionen

Relationenschemata (Relationsnamen, Attribute)

R, S Relationen der Relationenschemata \mathcal{R} , \mathcal{S}

A, B, C,... einfache Attribute

 $A = \{A_1,...,A_n\}$ Attributmenge eines Relationenschemas

W, X, Y, Z,... Mengen von Attributen

 $XY \equiv X \cup Y$ Mengen brauchen nicht disjunkt zu sein

a, b, c Werte einfacher Attribute

x, y, z Werte von X, Y, Z



FUNKTIONALE ABHÄNGIGKEIT

- Definition: funktionale Abhängigkeit (FA)
 - Die FA X → Y gilt (X bestimmt Y funktional), wenn für alle R von R gilt:
 - zwei Tupel, deren Komponenten in X übereinstimmen, stimmen auch in Y überein.
 - $\forall u, v \in R: u[X]) = v[X] \Rightarrow u[Y] = v[Y]$
 - alternativ: Relation R erfüllt die FA X \rightarrow Y, wenn für jeden X-Wert x der Ausdruck $\pi_Y(\sigma_{X=x}(R))$ höchstens ein Tupel hat.

	R						
Α	В	С	D				
a4	b2	с4	d3				
a1	b1	с1	d1				
a1	b1	с1	d2				
a2	b2	сЗ	d2				
аЗ	b2	с4	d3				

$$A \rightarrow B$$

$$C, D \rightarrow B$$

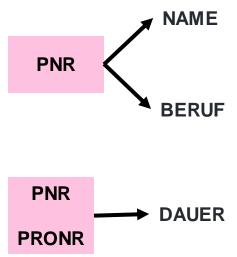
$$D \rightarrow C$$

$$B \rightarrow A$$



FUNKTIONALE ABHÄNGIGKEITEN (2)

– graphische Notation:



- FA beschreiben semantische Integritätsbedingungen bezüglich der Attribute eines Relationenschemas, die jederzeit erfüllt sein müssen
- triviale FA:
 - X → Y und Y ist Teilmenge von X
 - Spezialfall: $X \rightarrow X$



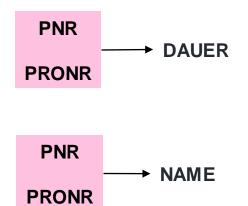
FUNKTIONALE ABHÄNGIGKEITEN (3)

Definitionen: voll funktionale vs. partielle Abhängigkeit

Sei
$$A_1, A_2, ..., A_n \rightarrow B_1, B_2, ..., B_m$$

 $B = \{B_1, B_2, ..., B_m\}$ ist **voll funktional abhängig** von $A = \{A_1, A_2, ..., A_n\}$, wenn B funktional abhängig von A ist, aber nicht funktional abhängig von einer echten Teilmenge von A ist.

 $A \rightarrow B$ ist eine *partielle Abhängigkeit*, wenn ein Attribut A_i in A existiert, so dass $(A - \{A_i\}) \rightarrow B$ gilt.



FUNKTIONALE ABHÄNGIGKEITEN: ÄQUIVALENZBEZIEHUNGEN



- Splitten / Kombinieren von FA
 - eine FA
 A1, A2, ..., An → B1, B2, ..., Bm
 - ist äquivalent zu m FA
 A1, A2, ..., An → B1

. . .

A1, A2, ..., An
$$\rightarrow$$
 Bm

 weitere Äquivalenzbeziehungen (Regeln zur Ableitung neuer aus gegebenen FA; Armstrong-Axiome)

R: Reflexivität: wenn X ⊆ Y

dann $Y \rightarrow X$ (triviale FA)

- K: Komplementierung: wenn $X \rightarrow Y$

dann $XZ \rightarrow YZ$

- T: Transitivität: wenn $X \rightarrow Y$, $Y \rightarrow Z$

 $dann X \rightarrow Z$

Beispiele

- R: PNR, Name →
- K: PNR, Name → Gehalt, Name wenn PNR → Gehalt
- T: PNR \rightarrow ANR, ANR \rightarrow ANAME: PNR \rightarrow ANAME



FA UND SCHLÜSSELKANDIDATEN

- X ist Schlüsselkandidat von R, wenn
 - für alle Y aus R: $X \rightarrow Y$
 - keine echte Teilmenge von X bestimmt funktional alle anderen Attribute Y aus R (Minimalität)
- Kenntnis aller nicht-trivialen FA ermöglicht Bestimmung der Schlüsselkandidaten

Beispiel 1: Attribute A, B, C, D mit

 $-B \rightarrow A$

SK: C, da C→A,D (Transitivität)

- $-B \rightarrow D$
- $-C \rightarrow B$

Beispiel 2: Attribute A, B, C, D mit

- $-A \rightarrow C$
- B,C → D SK: A,B , da AB \rightarrow CB (Kompl) und AB \rightarrow D(Trans.)



FUNKTIONALE ABHÄNGIGKEITEN: BEISPIEL

■ Beispieltabelle *Film*

Titel	Jahr	Dauer	FilmTyp	StudioName	Schauspieler
Star Wars	1977	124	Farbe	Fox	Carrie Fisher
Star Wars	1977	124	Farbe	Fox	Mark Hamill
Star Wars	1977	124	Farbe	Fox	Harrison Ford
Werk ohne Autor	2018	188	Farbe	Pergamon Film	Tom Schilling
Troja	2004	156	Farbe	Warner Bros	Brad Pitt
Troja	1956	118	SW	Warner Bros	Stanley Baker

funktionale Abhängigkeiten:

- Titel, Jahr \rightarrow Jahr
- Titel, Jahr → Dauer
- Titel, Jahr \rightarrow FilmTyp
- Schlüsselkandidat: Titel, Jahr, Schauspieler



"SCHLECHTE" RELATIONENSCHEMATA

	ProfVorl						
PersNr	Name	Fach	Raum	VorlNr	Titel	SWS	
3678	Rahm	DBS	356	5041	DBS1	3	
3678	Rahm	DBS	356	5049	DBS2	3	
3678	Rahm	DBS	356	4052	IDBS	4	
	•••		•••	•••	•••		
1234	Brewka	KI	152	5259	Wissensrepräsentation	2	
2137	Meyer	TI	17	4630	Informationstheorie	4	

- Update-Anomalien
 - Umzug von Raum 356 in Raum 338. Was passiert?
- Einfüge-Anomalien
 - neuer Prof ohne Vorlesungen?
- Löschanomalien
 - letzte Vorlesung eines Profs wird gelöscht?



ZERLEGUNG (DEKOMPOSITION) VON RELATIONEN

- 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 , ..., R_n der neuen Relationenschemata R_1 , ..., R_n rekonstruierbar sein.
 - erfordert vollständige Aufteilung der Attributmenge A von Schema \mathcal{R} , z.B. $A = A_1 \cup A_2$

$$R_1 := \pi_{A1} (R),$$

 $R_2 := \pi_{A2} (R)$

- Verlustfreiheit verlangt, dass für jede mögliche (gültige) Ausprägung R von $\mathcal R$ gilt:

$$R = R_1 \bowtie R_2$$

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



BEISPIEL: ZERLEGUNG VON PROFVORL

ProfVorI

PersNr	Name	Fach	Raum	<u>VorlNr</u>	Titel	SWS
3678	Rahm	DBS	356	5041	DBS1	3
3678	Rahm	DBS	356	5049	DBS2	3
3678	Rahm	DBS	356	4052	IDBS	4
	•••	•••				•••
1234	Brewka	KI	152	5259	Wissensrepräsentati on	2
2137	Meyer	TI	17	4630	Informationstheorie	4

Vorl

Prof

<u>PersNr</u>	Name	Fach	Raum
3678	Rahm	DBS	356
•••			
1234	Brewka	KI	152
2137	Meyer	TI	17

<u>VorINr</u>	PersNr(FK auf Prof)	Titel	SWS
5041	3678	DBS1	3
5049	3678	DBS2	3
4052	3678	IDBS	4
5259	1234	Wissensrepräsentation	2
4630	2137	Informationstheorie	4



"VERLUSTBEHAFTETE" ZERLEGUNG

Restaurant	Gast	Gericht
Firenze	Kowalski	Pizza
Roma	Meyer	Pizza
Roma	Kowalski	Calamari

Annahme (FA):

Restaurant, Gast -> Gericht

 $\pi_{RESTAURANT,\,GAST}$

Restaurant	Gast
Firenze	Kowalski
Roma	Meyer
Roma	Kowalski

 $\pi_{GAST,GERICHT}$

Gast	Gericht
Kowalski	Pizza
Meyer	Pizza
Kowalski	Calamari

Iggt

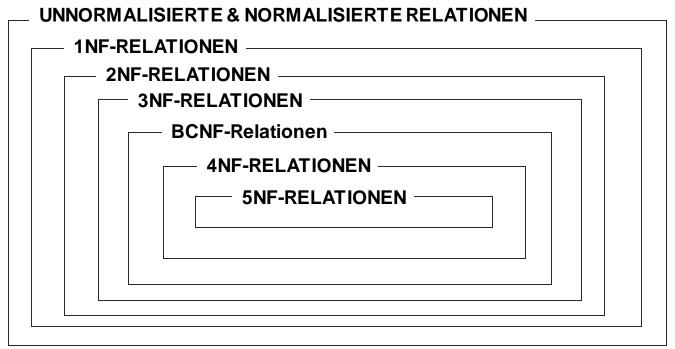
Besucht

Restaurant	Gast	Gericht
Firenze	Kowalski	Pizza
Firenze	Kowalski	Calamari
Roma	Meyer	Pizza
Roma	Kowalski	Pizza
Roma	Kowalski	Calamari

Funktionale Abhängigkeit wird durch Zerlegung aufgebrochen



NORMALISIERUNG VON RELATIONEN



- Zerlegung eines Relationenschemas R in höhere Normalformen
 - fortgesetzte Anwendung der Projektion im Zerlegungsprozess
 - Beseitigung von Anomalien bei Änderungsoperationen
 - Erhaltung aller nicht-redundanter Funktionalabhängigkeiten von \mathcal{R} (\rightarrow sie bestimmen den Informationsgehalt von \mathcal{R})
 - Gewährleistung der Rekonstruktion von R durch verlustfreie Verbünde
 - bessere "Lesbarkeit" der aus $\mathcal R$ gewonnenen Relationen



NORMALISIERUNG VON RELATIONEN (2)

<u>PNR</u>	PNAME	FACH	STUDENT (MATNR, NAME,)
3678	Rahm	DBS	196481 Maier
3678	Rahm	DBS	123766 Coy
3678	Rahm	DBS	900550 Schmitt
1234	Brewka	KI	654711 Abel .HH
1234	Brewka	KI	123766 Coy

Prüfungsgeschehen

Anomalien, z.B.:

- Insert Student
- **Delete** Prof
- Update Student
- Unnormalisierte Relation: Non-First Normal-Form (NF²)
 - enthält "Attribute", die wiederum Relationen sind (→ "geschachtelte" Relationen)
 - Darstellung von komplexen Objekten (hierarchische Sichten, Clusterbildung)
- Nachteile:
 - Unsymmetrie (nur eine Richtung der Beziehung)
 - implizite Darstellung von Information
 - Redundanzen bei (n:m)-Beziehungen
 - Anomalien bei Aktualisierung
- Normalisierung (ungünstige Lösung):
 - "Herunterkopieren" von Werten
 - Informationsgehalt wird erhalten
 - hoher Grad an Redundanz → Zerlegung von Relationen



ÜBERFÜHRUNG IN 1 NF

- Normalisierung (⇒ 1NF):
 - 1. Starte mit der übergeordneten Relation
 - 2. Nimm ihren Primärschlüssel und erweitere jede unmittelbar untergeordnete Relation damit zu einer selbständigen Relation.
 - 3. Streiche Attribute der untergeordnete Relationen aus der übergeordneten Relation.
 - 4. Wiederhole diesen Prozess ggf. rekursiv.

– Regeln:

- nicht-einfache Attribute bilden neue Relationen.
- Primärschlüssel der übergeordneten wird an untergeordnete Relation angehängt ('copy down the key')



ÜBERFÜHRUNG IN 1 NF

Prüfungsgeschehen (PNR, PNAME, FACH, STUDENT)

(MATNR, NAME, GEBORT, ADR, FNR, FNAME, DEKAN, PDAT, NOTE)

STUDENT = Wiederholungsgruppe mit 9 einfachen Attributen (untergeordnete Relation)

Relationenschema in 1NF

- PRÜFER (PNR, PNAME, FACH)
- PRÜFUNG (PNR (FK auf Prüfer), MATNR, NAME, GEBORT, ADR, FNR, FNAME, DEKAN, PDAT, NOTE)



ÜBERFÜHRUNG IN 2NF

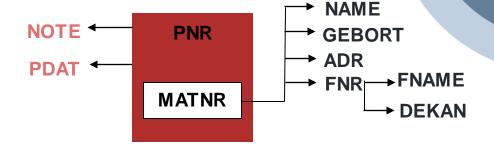
- 1NF verursacht immer noch viele Änderungsanomalien
 - verschiedene Entity-Mengen in einer Relation möglich bzw.
 Redundanz innerhalb einer Relation (Bsp.: PRÜFUNG)
- 2NF vermeidet einige Anomalien durch Eliminierung partiell abhängiger Attribute
 - Separierung verschiedener Entity-Mengen in eigene Relationen
- Def.: Primärattribut (Schlüsselattribut) Attribut, das zu mind. einem Schlüsselkandidaten eines Schemas gehört.
- Ein Relationenschema \mathcal{R} ist in 2NF, wenn es
 - in 1NF ist und
 - jedes **Nicht-Primärattribut** von $\mathcal R$ **voll funktional** von jedem Schlüsselkandidaten in $\mathcal R$ abhängt.
- Überführung in 2NF:
 - Bestimme funktionale Abhängigkeiten zwischen Nicht-Primärattributen und Schlüsselkandidaten
 - Eliminiere partiell abhängige Attribute und fasse sie in eigener Relation zusammen (unter Hinzunahme der zugehörigen Primärattribute)



ÜBERFÜHRUNG IN 2NF (2)

 voll funktionale Abhängigkeiten in PRÜFUNG

Relationenschema in 2NF



Prüfung⁴

PNR(FK auf Prüfer)	MATNR(FK auf Student)	PDAT	NOTE
1234	123 766	22.10.	4
1234	654711	14.02.	3
3678	196 481	21.09.	2
3678	123 766	02.03.	4
8223	226 302	12.07.	1

Prüfer

<u>PNR</u>	PNAME	FACH
1234	Brewka	KI
3678	Rahm	DBS
8223	Weber	WI

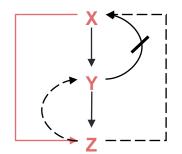
Student'

.6	MATNR	NAME	GEBORT	ADR	FNR	FNAME	DEKAN
	123 766	Coy	Leipzig	XX	F11	Wirtschaftswissenschaften	А
	654711	Abel	Torgau	XY	F19	Mathematik/Informatik	В
	196 481	Maier	Köln	YX	F19	Mathematik/Informatik	В
	226 302	Schulz	Leipzig	YY	F11	Wirtschaftswissenschaften	А



ÜBERFÜHRUNG IN 3NF

- Änderungsanomalien in 2NF sind immer noch möglich aufgrund von transitiven Abhängigkeiten.
- Beispiel: Vermischung von Fakultäts- und Studentendaten in Student'
 Definitionen:
- Eine Attributmenge Z von Relationenschema \mathcal{R} ist transitiv abhängig von einer Attributmenge X in R, wenn gilt:
- X und Z sind disjunkt
- es existiert eine Attributmenge Y in \mathcal{R} , so dass gilt: $X \to Y, Y \to Z, Y \nrightarrow X, Z \not\subseteq Y$



 $Z \rightarrow Y$ zulässig

strikte Transitivität: Z → Y

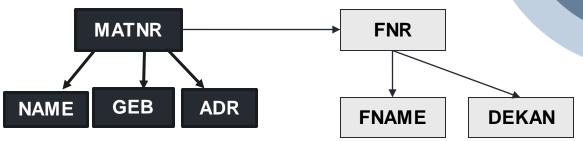
 Ein Relationenschema \mathcal{R} befindet sich in 3NF, wenn es sich in 2NF befindet und jedes
 Nicht-Primärattribut von \mathcal{R} von keinem Schlüsselkandidaten von \mathcal{R} transitiv

abhängig ist.



ÜBERFÜHRUNG IN 3NF (2)

funktionaleAbhängigkeiten in STUDENT'



Relationenschema in 3NF

Prüfung[•]

Fakultät

<u>FNR</u>	FNAME	DEKAN
F11	Wirtschaftswissenschaften	A
F12	Medizin	С
F19	Mathematik/Informatik	В

ľ	PNR(FK auf Prüfer)	MATNR(FK auf Student)	PDAT	NOTE
	1234	123 766	22.10.	4
	1234	654 711	14.02.	3
	3678	196 481	21.09.	2
	3678	123 766	02.03.	4
	8223	226 302	12.07.	1

Student

Prüfer

<u>PNR</u>	PNAME	FACH
1234	Brewka	KI
3678	Rahm	DBS
8223	Weber	WI

<u>MATNR</u>	NAME	GEBORT	ADR	FNR(FK auf Fakultät)
123 766	Coy	Leipzig	XX	F11
654711	Abel	Torgau	XY	F19
196 481	Maier	Köln	YX	F19
226 302	Schulz	Leipzig	YY	F11



BOYCE/CODD-NORMALFORM (BCNF)

- Definition der 3NF hat gewisse Schwächen bei Relationen mit mehreren, sich überlappenden Schlüsselkandidaten
- Beispiel:
- PRÜFUNG (PNR, MATNR, FACH, NOTE)
 PRIMARY KEY (PNR, MATNR),
 UNIQUE (MATNR, FACH)
 - es bestehe eine (1:1)-Beziehung zwischen
 PNR und FACH
 - einziges Nicht-Primärattribut: NOTE
 - ⇒ PRÜFUNG ist in 3NF
 - jedoch Änderungsanomalien, z. B. bei FACH
- Ziel: Beseitigung der Anomalien für Primärattribute
- Definition: Ein Attribut (oder eine Gruppe von Attributen), von dem andere voll funktional abhängen, heißt *Determinant*.
- welches sind die Determinanten in PRÜFUNG?

<u>PNR</u>	MATNR	Fach	NOTE
3678	196481	Datenbanksysteme	1
3678	123766	Datenbanksysteme	3
3678	900550	Datenbanksyteme	2
1234	654711	Künstliche Intelligenz	4



BOYCE/CODD-NORMALFORM (2)

 Definition: Ein Relationenschema R ist in BCNF, wenn es in 1NF ist und jeder Determinant ein Schlüsselkandidat von R ist.

Definition

- Ein Relationenschema ist in *BCNF*, falls gilt: Wenn eine Sammlung von Attributen Y (voll funktional) abhängt von einer disjunkten Sammlung von Attributen X, dann hängt jede andere Sammlung von Attributen Z auch von X (voll funktional) ab.
- D. h. für alle X, Y, Z mit X und Y disjunkt gilt: $X \rightarrow Y$ impliziert $X \rightarrow Z$
- Zerlegung von Prüfung

PRÜF (PNR, MATNR, NOTE) FBEZ (PNR, FACH) oder

PRÜF2 (MATNR, FACH, NOTE) FBEZ (PNR, FACH)

- beide Zerlegungen führen auf BCNF-Relationen
 - Änderungsanomalie ist verschwunden
 - alle funktionalen Abhängigkeiten sind erhalten



PROBLEME DER NORMALISIERUNG

sind BCNF-Zerlegungen immer sinnvoll?

Beispiel:

 $PRUFER(C) \rightarrow FACH(B),$ STUDENT, FACH $(A,B) \rightarrow PRÜFER(C)$ Prüfer Determinant, aber kein SK

 jeder Prüfer prüft nur ein Fach (aber ein Fach) kann von mehreren geprüft werden)

 jeder Student legt in einem bestimmten Fach nur Prüfungen bei einem Prüfer ab

– wie sieht BCNF-Zerlegung aus?

ist in 3NF, weil B Primärattribut ist!

SFP

STUDENT	<u>FACH</u>	PRÜFER
Sloppy	DBS	Rahm
Hazy	KI	Brewka
Sloppy	KI	Meier

PRÜFER	FACH
Rahm	DBS
Brewka	KI
Meier	КІ

STUDENT	PRÜFER
Sloppy	Rahm
Hazy	Brewka
Sloppy	Meier

– neue Probleme:

Hazv

Meier

- Abhängigkeit STUDENT, FACH → PRÜFER wird nicht erhalten
- BCNF hier zu streng, um bei der Zerlegung alle funktionalen Abhängigkeiten zu bewahren (key breaking dependency)



PROBLEME DER NORMALISIERUNG (2)

- weitestgehende Zerlegung nicht immer sinnvoll
- Beispiel:

```
Relation PERS (PNR, PLZ, ORT) mit FA PLZ \rightarrow ORT
```

Normalisierung verlangt Zerlegung in

```
PERS' (<u>PNR</u>, PLZ)
PZ (<u>PLZ</u>, Ort)
```

- Klärungsbedarf
 - Änderungshäufigkeit?
 - Suchaufwand für Adresse ? (Verbundoperation) !
 - sind ORT oder PLZ in diesem Kontext eigenständige Entities (als Kandidaten für eigene Relation in 3NF)?
- ⇒ besser PERS in 2NF!



ZUSAMMENFASSUNG

- Normalisierung von Relationen
 - Verbesserung eines gegebenen DB-Entwurfs
 - Ziel: eine Relation beschreibt nur einen Objekttyp
 - Eliminierung von Änderungsanomalien
 - wachsender Informationsgehalt mit zunehmender Normalisierung
- Bestimmung aller funktionalen Abhängigkeiten
 - n:1-Beziehung zwischen zwei Attributmengen einer Relation
 - wesentliche Integritätszusicherungen
- schrittweise Normalisierung:
 - 1NF: normalisierte Relationen (einfache Attribute)
 - 2NF: keine partiellen (funktionalen) Abhängigkeiten
 - 3NF: keine transitiven Abhängigkeiten (jedes Nicht-Primärattribut ist direkt von jedem SK abhängig)
 - BCNF: jeder Determinant ist Schlüsselkandidat
 - 3NF meist ausreichend
- Überarbeitung des DB-Schemas: Stabilitätsgesichtspunkte/Änderungshäufigkeiten können schwächere Normalformen verlangen

https://pingo.coactum.de/774240

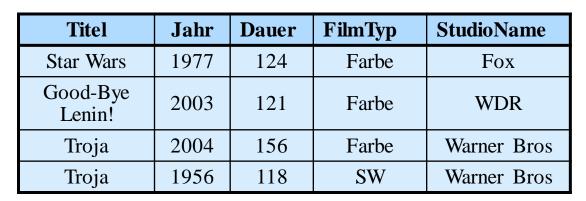
Normalisierung

Wählen Sie alle richtigen Antwortmöglichkeiten aus:	Wählen Sie alle richtigen Antwortmöglichkeiten aus:				
In der 2. Normalform werden transitive Abhängigkeiten ausgeschlossen					
Jedes Relationenschema in 3. Normalform ist auch in der 2. Normalform					
☐ Wenn A -> BC gilt, dann gilt auch A-> C					
Jeder Determinant ist ein Schlüsselkandidat					
☐ Das Relationenschema R(A,B,C) mit A->C, C->B ist in der 3. Normalform					
☐ In BCNF gibt es keine Determinanten, die nicht auch Schlüsselkandidaten sind					

Beispiel: Zerlegung der Filmtabelle

Titel	Jahr	Dauer	FilmTyp	StudioName	Schauspieler
Star Wars	1977	124	Farbe	Fox	Carrie Fisher
Star Wars	1977	124	Farbe	Fox	Mark Hamill
Star Wars	1977	124	Farbe	Fox	Harrison Ford
Good-Bye Lenin!	2003	121	Farbe	WDR	Daniel Brühl
Troja	2004	156	Farbe	Warner Bros	Brad Pitt
Troja	1956	118	SW	Warner Bros	Stanley Baker







Titel	Jahr	Schauspieler
Star Wars	1977	Carrie Fisher
Star Wars	1977	Mark Hamill
Star Wars	1977	Harrison Ford
Good-Bye Lenin!	2003	Daniel Brühl
Troja	2004	Brad Pitt
Troja	1956	Stanley Baker

© Prof. Dr. E. Rahm 6 - 30