Datenmodellierung/Datenbanksysteme

VU 184.685/VU 184.686, WS 2022

Relationale Entwurfstheorie – Normalformen

Johannes Fichte, Felix Winter

Institut für Logic and Computation
Technische Universität Wien

FAKULTÄT
FÜR !NFORMATIK

Faculty of Informatics

Acknowledgments

Die Folien sind eine Weiterentwicklung der Folien von Katrin Seyr und Sebastian Skritek, welche wiederum auf den zum Lehrbuch zur Verfügung gestellten Folien von A. Kemper basieren.

Der Inhalt basiert auf und behandelt Kapitel 6 des Lehrbuchs (Kemper, Eickler: Datenbanksysteme – Eine Einführung)

Überblick Wh.

- Ziele
- Funktionale Abhängigkeiten
  - Definitionen
  - Kanonische Überdeckung
- Entwurfstheorie und Zerlegung
  - "Schlechte" Relationenschemata
  - Zerlegung von Relationenschemata
  - Kriterien für eine "sinnvolle" Zerlegung
- Normalformen (1., 2., 3., Boyce-Codd)
  - Normalisierung durch Synthesealgorithmus
  - Normalisierung durch Dekomposition

Entwurfstheorie und Zerlegung

- "Schlechte" Relationenschemata
- Zerlegung von Relationenschemata
- Kriterien für eine "sinnvolle" Zerlegung
  - Verlustlosigkeit
  - Abhängigkeitstreue

"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	5295	Der Wiener Kreis	2
2137	Kant	C4	7	4630	Die drei Kritiken	4

Update-Anomalie: Sokrates zieht um

Lösch-Anomalie: Bsp: "Die 3 Kritiken" fällt weg

Einfügeanomalie: Bsp: Curie ist neu und liest noch keine Vorlesungen (hat keinen Schlüssel?!)

Zerlegung von Relationenschemata

- Anomalien beruhen auf der Tatsache, dass nicht zusammenpassende Informationen zusammen gespeichert wurden.
- Intuitiv:
  - alle Informationen zu Professoren werden in einer Relation gespeichert,
  - alle Informationen zu Vorlesungen werden in einer Relation gespeichert,
  - die Information über die Verknüpfung beider Relationen wird in einer Relation gespeichert.
- Lösung: Zerlegung des Schemas in Teilschemata
- Frage: was ist eine "sinnvolle" Zerlegung?

Zerlegung von Relationenschemata

Beispiel

Zerlegung von

ProfVorl(PersNr, Name, Rang, Raum, VorlNr, Titel, SWS)

in Teilschemata

Versuch 1: Zerlegung in:

- Prof(PersNr, Name, Rang, Raum)
- Vorlesung(VorlNr, Titel, SWS)

Problem: Wer hält welche LVA?

Ursache: Teilschemata können nicht mehr korrekt verknüpft werden.

Zerlegung von Relationenschemata

Beispiel

Zerlegung von

ProfVorl(PersNr, Name, Rang, Raum, VorlNr, Titel, SWS)

in Teilschemata

Versuch 2: Zerlegung in:

- Prof(PersNr, Name, Raum, VorlNr, Titel, SWS)
- Titel(Name, Rang)

Problem: Rang von Professoren gleichen Namens.

Ursache: Die FD $\{\text{PersNr}\} \rightarrow \{\text{Rang}\}$ ist verlorengegangen.

Zerlegung von Relationenschemata – Korrektheit

- Eine Zerlegung eines Relationenschemas $\mathcal{R}$ ist eine Menge von Schemata $\mathcal{R}_1, \ldots, \mathcal{R}_n$ so dass gilt:

$$att(\mathcal{R}_1) \cup \cdots \cup att(\mathcal{R}_n) = att(\mathcal{R})$$

- Es gibt zwei Korrektheitskriterien für die Zerlegung von Relationenschemata:

  Verlustlosigkeit (Verlust von Information)
  Abhängigkeitstreue (Verlust von Metainformation)

Zerlegung von Relationenschemata – Korrektheit

Definition (Verlustlosigkeit - intuitiv)

Die in der Ausprägung R des Schemas $\mathcal{R}$ enthaltenen Informationen müssen aus den Ausprägungen $R_1, \ldots, R_n$ der neuen Schemata $\mathcal{R}_1, \ldots, \mathcal{R}_n$ rekonstruierbar sein.

Definition (Abhängigkeitstreue - intuitiv)

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

Zerlegung von Relationenschemata – Verlustlosigkeit

Definition (Verlustlosigkeit)

Gegeben ein Relationenschema $(\mathcal{R}, F_{\mathcal{R}})$.
Eine Zerlegung $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_n$ von $\mathcal{R}$ ist verlustlos, falls für jede Ausprägung R von $\mathcal{R}$ welche $F_{\mathcal{R}}$ erfüllt gilt:

$$R = \pi_{\mathcal{R}_1}(R) \bowtie \pi_{\mathcal{R}_2}(R) \bowtie \cdots \bowtie \pi_{\mathcal{R}_n}(R)$$

Theorem (Hinreichende Bedingung für Verlustlosigkeit)

Eine Zerlegung von $\mathcal{R}$ in $\mathcal{R}_1$ und $\mathcal{R}_2$ ist verlustlos, wenn die Joinattribute in einer der Teilrelationen (Super-) Schlüssel sind.

Zerlegung von Relationenschemata – Verlustlosigkeit

Beispiel (Verlustlose Zerlegung 1)

Zerlegung von ProfVorl(PersNr, Name, Rang, VorlNr, Titel, SWS) in mehrere Teilschemata:

ProfVorl

PersNr	Name	Rang	VorlNr	Titel	SWS
2125	Sokrates	C4	5041	Ethik	4
2125	Sokrates	C4	5049	Mäeutik	2
2125	Sokrates	C4	4052	Logik	4
...	...	...	...	...	...
2132	Popper	C3	5295	Der Wiener Kreis	2
2137	Kant	C4	4630	Die drei Kritiken	4

Zerlegung von Relationenschemata – Verlustlosigkeit

Beispiel (Verlustlose Zerlegung 1)

Versuch 1:

Professoren

PersNr	Name	Rang
2125	Sokrates	C4
2132	Popper	C3
2137	Kant	C4
...	...	...

Vorlesungen

VorlNr	Titel	SWS
5041	Ethik	4
5049	Mäeutik	2
4052	Logik	4
5295	Der Wiener Kreis	2
4630	Die drei Kritiken	4
...	...	...

Frage: gilt ProfVorl = Professoren ⋈ Vorlesungen ?

Nein: die Verknüpfung mittels Join wird mangels Joinattributs zum Karthesischen Produkt.

Zerlegung von Relationenschemata – Verlustlosigkeit

Beispiel (Verlustlose Zerlegung 1)

Versuch 2:

Professoren

PersNr	Name	Rang
2125	Sokrates	C4
2132	Popper	C3
2137	Kant	C4
...	...	...

Vorlesungen

VorlNr	Titel	SWS	PersNr
5041	Ethik	4	2125
5049	Mäeutik	2	2125
4052	Logik	4	2125
5295	Der Wiener Kreis	2	2132
4630	Die drei Kritiken	4	2137
...	...	...	...

Frage: gilt ProfVorl = Professoren $\bowtie$ Vorlesungen ?

Ja: das Joinattribut PersNr ist Schlüssel in Professoren $\Rightarrow$ hinreichende Bedingung ist erfüllt
$\Rightarrow$ Zerlegung ist verlustlos

Zerlegung von Relationenschemata – Verlustlosigkeit

Beispiel (Verlustlose Zerlegung 2)

Zerlege Relation Eltern(Vater, Mutter, <u>Kind</u>) in zwei Teilrelationen:

$\mathcal{R}_1$ Väter(Vater, <u>Kind</u>)

$\mathcal{R}_2$ Mütter(Mutter, <u>Kind</u>)

Frage: ist die Zerlegung verlustlos?

Ja: das Joinattribut von Väter mit Mütter ist Kind
$\Rightarrow$ Verlustlosigkeit ist garantiert

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

Zerlegung von Relationenschemata – Abhängigkeitstreue

- Sei F eine Menge an FDs über $\mathcal{R}$, und $\mathcal{R}' \subseteq \mathcal{R}$:

$$F[\mathcal{R}'] = \{\alpha \to \beta \in F \mid \alpha \cup \beta \subseteq \mathcal{R}'\}$$

Definition (Abhängigkeitstreue)

Gegeben ein Relationenschema $(\mathcal{R}, F)$.
Eine Zerlegung $\mathcal{R}_1, \ldots, \mathcal{R}_n$ von $\mathcal{R}$ ist abhängigkeitstreu, wenn

$$F \equiv (F^+[\mathcal{R}_1] \cup \cdots \cup F^+[\mathcal{R}_n]) \text{ dh. } F^+ = (F^+[\mathcal{R}_1] \cup \cdots \cup F^+[\mathcal{R}_n])^+$$

Alternativ: abhängigkeitstreu wenn es eine Menge G von FDs gibt, so dass $G \equiv F$ und $F \equiv (G[\mathcal{R}_1] \cup \cdots \cup G[\mathcal{R}_n])$

Zerlegung von Relationenschemata – Abhängigkeitstreue

Beispiel (Keine Abhängigkeitstreue Zerlegung)

Gegeben ein Relationenschema Adresse(Straße, Ort, BLand, PLZ) und folgenden Annahmen:

- Orte werden durch Namen und BLand eindeutig identifiziert.
- PLZgebiete gehen nicht über Ortsgrenzen hinweg $\Rightarrow$ $\{\text{PLZ}\} \rightarrow \{\text{Ort, BLand}\}$
- Innerhalb einer Straße ändert sich die PLZ nicht $\Rightarrow$ $\{\text{Straße, Ort, BLand}\} \rightarrow \{\text{PLZ}\}$.

Zerlegung in:

$\mathcal{R}_1$ Straßen(<u>PLZ, Straße</u>)

$\mathcal{R}_2$ Orte(<u>PLZ</u>, Ort, BLand) mit FD $\{\text{PLZ}\} \rightarrow \{\text{Ort, BLand}\}$

ABER: $\{\text{Straße, Ort, BLand}\} \rightarrow \{\text{PLZ}\}$ ist verloren gegangen.
(Zerlegung ist verlustlos, da PLZ Schlüssel in Orte)

Zerlegung von Relationenschemata – Abhängigkeitstreue

Beispiel (Keine Abhängigkeitstreue Zerlegung)

Adresse

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

Straßen

PLZ	Straße
60313	Goethestraße
60437	Schillerstraße
15234	Goethestraße
15235	Goethestraße

Orte

Ort	BLand	PLZ
Frankfurt	Hessen	60313
Frankfurt	Hessen	60437
Frankfurt	Brandenburg	15234
Frankfurt	Brandenburg	15235

Zerlegung von Relationenschemata – Abhängigkeitstreue

Beispiel (Warum $F^+[\mathcal{R}_i]$?)

Gegeben ein Relationenschema $(\mathcal{R}, F)$ mit

$$\begin{aligned} \mathcal{R} &= \text{Vorfahren(Kind, Mutter, Oma) und} \\ F &= \{\{\text{Kind}\} \rightarrow \{\text{Mutter}\}, \{\text{Kind, Mutter}\} \rightarrow \{\text{Oma}\}\} \end{aligned}$$

Zerlegung in

$\mathcal{R}_1$ Mütter(Kind, Mutter), $F[\mathcal{R}_1] = \{\{\text{Kind}\} \rightarrow \{\text{Mutter}\}\}$
$F^+[\mathcal{R}_1] = \{\{\text{Kind}\} \rightarrow \{\text{Mutter}\}, \dots\}$

$\mathcal{R}_2$ Omas(Kind, Oma), $F[\mathcal{R}_2] = \emptyset$
$F^+[\mathcal{R}_2] = \{\{\text{Kind}\} \rightarrow \{\text{Oma}\}, \dots\}$

$(F[\mathcal{R}_1] \cup F[\mathcal{R}_2])^+ = \{\{\text{Kind}\} \rightarrow \{\text{Mutter}\}, \dots\} \neq F^+$
$(F^+[\mathcal{R}_1] \cup F^+[\mathcal{R}_2])^+ = \{\{\text{Kind}\} \rightarrow \{\text{Mutter}\}, \{\text{Kind}\} \rightarrow \{\text{Oma}\}, \{\text{Kind, Mutter}\} \rightarrow \{\text{Oma}\}, \dots\} = F^+$

Normalformen

- 1. Normalform
  - NF^2 Relationen
- 2. Normalform
- 3. Normalform
  - Zerlegung in 3NF
  - Synthesealgorithmus
- Boyce-Codd Normalform
  - Zerlegung in BCNF
  - Dekompositionsalgorithmus

1. Normalform

Definition (1. Normalform)

Ein Relationenschema $\mathcal{R}$ ist in 1. Normalform, wenn die Domänen von $\mathcal{R}$ atomar sind.

Eltern		
Vater	Mutter	Kinder
Johann	Martha	{Else, Lucie}
Heinz	Martha	{Cleo}

nicht in 1NF

Eltern		
Vater	Mutter	Kind
Johann	Martha	Else
Johann	Martha	Lucie
Heinz	Martha	Cleo

in 1NF

NF^2 Relationen

- Non-First Normal Form Relationen
- Attribut einer Relation kann selbst eine Relation sein
- im SQL2003 Standard enthalten

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

2. Normalform

- Intuitiv: Relationenschema $\mathcal{R}$ verletzt die 2. Normalform, wenn in der Relation Informationen über mehr als ein Konzept modelliert werden.
- Nur von historischem Interesse, da es immer noch zu Anomalien kommen kann.

3. Normalform

Definition (3. Normalform)

Ein Relationenschema $\mathcal{R}$ ist in dritter Normalform, wenn für jede auf $\mathcal{R}$ geltende funktionale Abhängigkeit der Form $\alpha \rightarrow B$ mit $\alpha \subseteq \mathcal{R}$ und $B \in \mathcal{R}$ mindestens eine der folgenden drei Bedingungen gilt:

1. $B \in \alpha$, d.h., die FD ist trivial
2. α ist Superschlüssel von $\mathcal{R}$.
3. das Attribut B ist in einem der Schlüssel von $\mathcal{R}$ enthalten

3. Normalform – "Alternative" Definition

Originaldefinition von Codd (1971):

"Ein Relationenschema befindet sich in 3NF wenn es sich in der 2NF befindet und kein Nichtschlüsselattribut transitiv von einem Schlüssel abhängt."

D.h. die folgenden Bedingungen müssen erfüllt sein:

- 2NF
- Es existieren keine Attributmengen α (Schlüssel), β (kein Schlüssel) und Attribut A (nicht in einem Schlüssel enthalten), so dass
  - $\alpha \rightarrow \beta$ und $\beta \rightarrow A$ gilt sowie
  - $\beta \not\rightarrow \alpha$, $A \notin \alpha$, $A \notin \beta$

"Every non-key attribute must provide a fact about the key, the whole key, and nothing but the key. (So help me Codd)"

3. Normalform

Beispiel

ProfVorl(PersNr, Name, Rang, Raum, VorlNr, Titel, SWS)

$F_d = \{$PersNr $\rightarrow$ Name Rang Raum,
VorlNr $\rightarrow$ PersNr Name Rang Raum VorlNr Titel SWS$\}$

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	5295	Der Wiener Kreis	2
2137	Kant	C4	7	4630	Die drei Kritiken	4

Nicht in 3. Normalform

3. Normalform

Beispiel (Check auf 3NF 1)

$$\mathcal{R} = (ABCDEF),$$
$$F = \{C \to BDAE\} = \{C \to B, C \to D, C \to A, C \to E\}.$$

Der einzige Schlüssel von $\mathcal{R}$ ist: CF

$(\mathcal{R}, F)$ in 3NF, wenn für jede FD eine der NF-Bedingungen gilt:

1. Ist die FD trivial? Nein, Bedingung 1 gilt für keine FD.
2. Ist die Attributmenge auf der linken Seite der FD ein Superschlüssel von $\mathcal{R}$? Nein, Bedingung 2 gilt für keine FD.
3. Ist das Attribut auf der rechten Seite der FD in einem der Schlüssel von $\mathcal{R}$ enthalten? Nein, Bdg. 3 gilt für keine FD.

Mindestens eine FD verletzt die 3NF $\Rightarrow \mathcal{R}$ ist nicht in 3NF.

3. Normalform

Beispiel (Check auf 3NF 2)

$$\mathcal{R} = (ABCDEF)$$
$$F = \{C \to B, C \to D, D \to A, D \to E, E \to C, E \to F, F \to E\}$$

Die Schlüssel von $\mathcal{R}$ sind: C, E, F, D

1. Ist die FD trivial? Nein, Bedingung 1 gilt für keine FD.
2. Ist die Attributmenge auf der linker Seite der FD ein Superschlüssel von $\mathcal{R}$? OK für $C \to B, C \to D, D \to A, D \to E, E \to C, E \to F, F \to E$.
3. Ist das Attribut auf der rechten Seite der FD in einem der Schlüssel von $\mathcal{R}$ enthalten? Nein für $C \to B, D \to A$ aber OK für $C \to D, D \to E, E \to CF, F \to E$.

Keine FD verletzt die 3NF $\Rightarrow$ in 3NF
(In diesem Fall schon nach Schritt 2 geklärt.)

3. Normalform

Beispiel

ProfVorl(PersNr, Name, Rang, Raum, VorlNr, Titel, SWS)

$F_d = \{$PersNr $\rightarrow$ Name Rang Raum,
VorlNr $\rightarrow$ PersNr Name Rang Raum VorlNr Titel SWS$\}$

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	5295	Der Wiener Kreis	2
2137	Kant	C4	7	4630	Die drei Kritiken	4

Nicht in 3. Normalform

Zerlegung in 3NF

Theorem

Ein Relationenschema $\mathcal{R} = \mathcal{R}_1 \cup \cdots \cup \mathcal{R}_n$ ist in dritter Normalform, wenn alle $\mathcal{R}_i$ in dritter Normalform sind.

- Gegeben Relationenschema $\mathcal{R}$ mit FDs F.
- Gesucht: Zerlegung in Teilschemata $\mathcal{R}_1 \ldots \mathcal{R}_n$, für die gilt:
  - Zerlegung in $\mathcal{R}_1 \ldots \mathcal{R}_n$ ist verlustlos,
  - Zerlegung in $\mathcal{R}_1 \ldots \mathcal{R}_n$ ist abhängigkeitstreu,
  - alle $\mathcal{R}_1 \ldots \mathcal{R}_n$ sind in dritter Normalform.
- Lösung: Synthesealgorithmus

Der Synthesealgorithmus

1. Bestimme kanonische Überdeckung F_c zu F.
2. Für jede funktionale Abhängigkeit $\alpha \rightarrow \beta \in F_c$:
   - Erstelle ein Relationenschema $\mathcal{R}_i := \alpha \cup \beta$
   - Ordne $\mathcal{R}_\alpha$ die FDs $F_i := F_c[\mathcal{R}_i]$ zu.

     Informell: aus jeder FD wird ein eigenes Schema (Bedingung für Abhängigkeitstreue erfüllt) - kanonische Überdeckung wichtig, um nicht zu viele oder zu große Teilschemata zu generieren
3. Enthält eines der in Schritt 2. erzeugten Teilschemata einen Schlüssel von $\mathcal{R}$ bzgl. $F_c \Rightarrow$ Schritt 4
   andernfalls: wähle einen Schlüssel $\kappa \in \mathcal{R}$ aus und definiere folgendes zusätzliche Schema: $\mathcal{R}_\kappa := \kappa$ mit $F_\kappa := \emptyset$

   Informell: erzeuge ein Schema zum Verknüpfen der Teilschemata (Bedingung für Verlustfreiheit erfüllt)
4. Eliminiere die in einem anderen Schema $\mathcal{R}_j$ enthaltenen Schemata $\mathcal{R}_i$.

   Informell: kürze überflüssige Schemata

Der Synthesealgorithmus

Beispiel (Synthesealgorithmus 1)

$\mathcal{R} = \{PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, BLand, EW, Landesregierung\} = \{P, N, R, Z, O, S, Plz, V, B, E, L\}$ und die FDs von früher.

1. Kanonische Überdeckung:
   $P \to NRZOSB$, $Z \to P$, $SBO \to Plz$, $OB \to EV$, $B \to L$, $Plz \to BO$
2. Generieren der Teilschemata und Zuordnung aller FDs:
   $\{PNRZOSB\}$ es gelten: $P \to NRZOSB$ und $Z \to P$
   $\{ZP\}$ es gilt: $Z \to P$
   $\{SBOPlz\}$ es gelten: $SBO \to Plz$ und $Plz \to BO$
   $\{OBEV\}$ es gilt: $OB \to EV$
   $\{BL\}$ es gilt: $B \to L$
   $\{PlzBO\}$ es gilt: $Plz \to BO$

Der Synthesealgorithmus

Beispiel (Synthesealgorithmus 1)

3. Enthält eines der Teilschemata einen Schlüssel von $\mathcal{R}$ bezüglich F?
   Ja: P war Schlüssel und ist in $\{PNRZOSB\}$ enthalten $\Rightarrow$ fertig
4. Schemaelimination:

   $\{ZP\}$ ist schon in $\{PNRZOSB\}$ enthalten $\Rightarrow$ kürzen

   $\{PlzBO\}$ ist schon in $\{SBOPlz\}$ enthalten $\Rightarrow$ kürzen

5. Ergebnis:

   Professoren: {[PersNr, Name, Rang, Raum, Ort, Straße, BLand]}
   PLZListe: {[Straße, Ort, BLand,PLZ]}
   Städte: {[Ort, BLand, EW, Vorwahl]}
   Regierung: {[BLand, Landesregierung]}

Der Synthesealgorithmus

Beispiel (Synthesealgorithmus 2)

$\mathcal{R} = (ABCDEF), F = \{A \to EC, BC \to F, D \to B\}$
Frage: in 3 NF? Nein. Schlüssel: AD

1. ist schon kanonisch
2. $\mathcal{R}_1 = (AEC), \mathcal{R}_2 = (BCF), \mathcal{R}_3 = (DB)$
   keine (nichttrivialen) Zuordnungen möglich
3. Schlüssel von $\mathcal{R}$: AD hinzufügen von $\mathcal{R}_4 = (AD)$
4. Nichts ist zu eliminieren.

$$\mathcal{R} = (AEC) \cup (BCF) \cup (DB) \cup (AD)$$

Der Synthesealgorithmus

Beispiel

Warum ist eine kanonische Überdeckung notwendig?

ProfVorl(PersNr, Name, Rang, Raum, VorlNr, Titel, SWS)

F_1: {PersNr → Name Rang Raum,
VorlNr → PersNr Name Rang Raum VorlNr Titel SWS}

F_2: {PersNr → Name, PersNr → Rang, PersNr → Raum,
VorlNr → PersNr, VorlNr → Name, VorlNr → Rang,
VorlNr → Raum, VorlNr → VorlNr, VorlNr → Titel,
VorlNr → SWS}

Ist 3 NF genug?

Beispiel

$$\mathcal{R} = \text{Städte(Ort, Land, Chef, EW)}$$

$$F = \{\{\text{Ort, Land}\} \rightarrow \{\text{EW}\}, \{\text{Land}\} \rightarrow \{\text{Chef}\}, \{\text{Chef}\} \rightarrow \{\text{Land}\}\}$$

Schlüssel: {Ort, Land}, {Ort, Chef} ⇒ in 3NF

Städte

Ort	Land	Chef	EW
Minas Tirith	Gondor	Denethor	2002
Osgiliath	Gondor	Denethor	0
Dol Amroth	Gondor	Denethor	700
Barad-dûr	Mordor	Sauron	20

Boyce-Codd Normalform

Definition (Boyce-Codd Normalform)

Ein Relationenschema $\mathcal{R}$ ist in Boyce-Codd Normalform, wenn für jede auf $\mathcal{R}$ geltende funktionale Abhängigkeit der Form $\alpha \rightarrow B$ mit $\alpha \subseteq \mathcal{R}$ und $B \in \mathcal{R}$ mindestens eine der folgenden zwei Bedingungen gilt:

1. $B \in \alpha$, d.h., die FD ist trivial
2. α ist Superschlüssel von $\mathcal{R}$.

Theorem (Normalformen Zusammenhang)

Die Normalformen hängen wie folgt zusammen:

$$BCNF \subseteq 3NF \subseteq 2NF \subseteq 1NF$$

Boyce-Codd Normalform

Beispiel (Check auf BCNF 1)

$$\mathcal{R} = (ABCDEF)$$
$$F = \{C \to B, C \to D, D \to A, D \to E, E \to C, E \to F, F \to E\}$$

Die Schlüssel von $\mathcal{R}$ sind: C, E, D, F
$(\mathcal{R}, F)$ ist in BCNF, wenn für jede FD eine der zwei NF-Bedingungen gilt:

1. Keine FD ist trivial: Bedingung 1 gilt nicht
2. Ist die Attributmenge auf der linken Seite der FDs ein Superschlüssel von $\mathcal{R}$? Ja, Bedingung 2 gilt für alle FDs.

Keine FD verletzt die BCNF $\Rightarrow$ in BCNF

Boyce-Codd Normalform

Beispiel (Check auf BCNF 2)

$\mathcal{R}$ = Städte(Ort, Land, Chef, EW)

F = {{Ort, Land} → {EW},
{Land} → {Chef}, {Chef} → {Land}}

Schlüssel: {Ort, Land} bzw. {Ort, Chef}

1. keine FD ist trivial
2. ist die Attributmenge auf der linken Seite der FDs ein Superschlüssel? Nein für Land → Chef, Chef → Land

⇒ $\mathcal{R}$ nicht in BCNF

Aber: in 3NF, da Bedingung 2 (3NF): auf der rechten Seite steht ein Schlüsselattribut erfüllt ist.

Zerlegung in BCNF

Theorem

Ein Relationenschema $\mathcal{R} = \mathcal{R}_1 \cup \cdots \cup \mathcal{R}_n$ ist in Boyce-Codd Normalform, wenn alle $\mathcal{R}_i$ in Boyce-Codd Normalform sind.

- Gegeben Relationenschema $\mathcal{R}$ mit FDs F.
- Gesucht: Zerlegung in Teilschemata $\mathcal{R}_1 \ldots \mathcal{R}_n$, für die gilt:
  - Zerlegung in $\mathcal{R}_1 \ldots \mathcal{R}_n$ ist verlustlos,
  - Zerlegung in $\mathcal{R}_1 \ldots \mathcal{R}_n$ ist abhängigkeitstreu,
  - alle $\mathcal{R}_1 \ldots \mathcal{R}_n$ sind in Boyce-Codd Normalform.
- nur die BCNF garantiert völlige Anomaliefreiheit
- Problem: eine verlustlose Zerlegung in BCNF ist immer möglich, aber die Abhängigkeitstreue ist nicht immer erzielbar
- Dekompositionsalgorithmus (ohne Abhängigkeitstreue)

Dekompositionsalgorithmus

- Starte mit $Z = \{(\mathcal{R}, F)\}$
- Solange es ein Relationenschema $(\mathcal{R}_i, F_i)$ in Z gibt, das die BCNF verletzt d.h. es gibt eine FD $\alpha \rightarrow \beta$ in $\mathcal{R}_i$ mit $\beta \not\subseteq \alpha$ und $\neg(\alpha \rightarrow \mathcal{R}_i)$:
  Wähle eine solche FD $\alpha \rightarrow \beta$ aus und zerlege wie folgt:
  - $\mathcal{R}_{i_1} := (\alpha \cup \beta)$, $F_{i_1} := F_i^+[\mathcal{R}_{i_1}]$
  - $\mathcal{R}_{i_2} := \mathcal{R}_i - (\beta - \alpha)$, $F_{i_2} := F_i^+[\mathcal{R}_{i_2}]$

  Entferne $(\mathcal{R}_i, F_i)$ aus Z und füge $(\mathcal{R}_{i_1}, F_{i_1})$ und $(\mathcal{R}_{i_2}, F_{i_2})$ ein:

$$Z := (Z - (\{\mathcal{R}_i, F_i)\}) \cup \{(\mathcal{R}_{i_1}, F_{i_1})\} \cup \{(\mathcal{R}_{i_2}, F_{i_2})\}$$

$$\mathcal{R}_i$$

$$\alpha \rightarrow \beta \text{ und weitere FDs } F_i$$

$$\mathcal{R}_{i_1} = (\alpha \cup \beta) \qquad \mathcal{R}_{i_2} = (\mathcal{R}_i - (\beta - \alpha))$$

$$F_{i_1} = F_i^+[\mathcal{R}_{i_1}] \qquad F_{i_2} = F_i^+[\mathcal{R}_{i_2}]$$

Dekompositionsalgorithmus

Beispiel (Dekompositionsalgorithmus 1)

$\mathcal{R}$ = Städte(Ort, Land, Chef, EW)

F = {{Ort, Land} → {EW},
{Land} → {Chef}, {Chef} → {Land}}

Schlüssel: {Ort, Land}, {Ort, Chef}

- $Z := \mathcal{R}$
- {Land} → {Chef} verletzt die BCNF
  - $\mathcal{R}_1$:= (Land, Chef) und
    $F_1 := \{\{Land\} \rightarrow \{Chef\}, \{Chef\} \rightarrow \{Land\}, \ldots\}$
  - $\mathcal{R}_2$:= (Ort, Land, EW) und $F_2 := \{\{Ort, Land\} \rightarrow \{EW\}, \ldots\}$
  - $Z := \{(\mathcal{R}_1, F_1), (\mathcal{R}_2, F_2)\}$
- die Zerlegung ist verlustfrei und abhängigkeitstreu

Dekompositionsalgorithmus

Beispiel (Dekompositionsalgorithmus 2)

$\mathcal{R}$ = PLZListe(Straße, Ort, BLand, PLZ)

F = {{PLZ}→{Ort, BLand}, {Straße, Ort, BLand}→{PLZ}}

Schlüssel: {Straße, Ort, BLand}, {PLZ, Straße},

- $Z := (\mathcal{R}, F)$
- {{PLZ} → {Ort, BLand}} verletzt die BCNF
  - $\mathcal{R}_1$:= (PLZ, Ort, BLand) und $F_1 := \{\{PLZ\} \rightarrow \{Ort, BLand\}, \dots\}$
  - $\mathcal{R}_2$:= (PLZ, Straße) und F_2 enthält nur triviale FDs
  - $Z := \{(\mathcal{R}_1, F_1), (\mathcal{R}_2, F_2)\}$
- die Zerlegung ist verlustfrei aber die Abhängigkeit {{Straße, Ort, BLand} → {PLZ}} ist verloren gegangen

Dekompositionsalgorithmus

Beispiel (Dekompositionsalgorithmus 3)

$$\mathcal{R} = ABCDE$$
$$F = \{A \to B, C \to B, BE \to D\}$$
Schlüssel: ACE

$$\mathcal{R}_1 = AB$$
$$F_1 = \{A \to B, \dots\}$$
Schlüssel: A

$$\mathcal{R}_2 = ACED$$
$$F_2 = \{AE \to D, CE \to D, ACE \to D, \dots\}$$
Schlüssel: ACE

$$\mathcal{R}_{2,1} = AED$$
$$F_{2,1} = \{AE \to D, \dots\}$$
Schlüssel: AE

$$\mathcal{R}_{2,2} = ACE$$
$$F_{2,2} = \{\dots\}$$
Schlüssel: ACE

Verlustfrei, aber nicht Abhängigkeitstreu:
$\{C \to B, BE \to D\} \in F$, z.B. $\{CE \to D, \dots\} \in F^+$

Lernziele

- Was für Arten von Anomalien gibt es?
- Was ist eine verlustlose/abhängigkeitstreue Zerlegung?
- Welche Normalformen gibt es?
  - Wann sind sie erfüllt, und wie kann man sie berechnen?