

Normalformen

Prime Attribute

Sei R ein Relationenschema. Ein Attribut $A \in R$ heißt *prim*, falls A Teil eines Schlüsselkandidaten von R ist. Andernfalls heißt A *nichtprim*.¹

Lernkarteikarte: Normalformen

1NF alle Attribute atomar

2NF 1NF + kein Nichtprimärattribut hängt funktional von einer echten Teilmenge eines Schlüsselkandidaten ab

3NF 2NF + keine transitiven Abhängigkeiten über Nichtschlüsselkandidaten

BCNF 3NF + jede Determinante ein Schlüsselkandidat^a

^aKemper und Eickler, *Datenbanksysteme*, Seite 179.

2NF verletzt

FA = {
 { A } \rightarrow { C },
 { B } \rightarrow { C },
}

3NF verletzt

FA = {
 { A } \rightarrow { BC },
 { B } \rightarrow { C },
}

BCNF verletzt

FA = {
 { AB } \rightarrow { C },
 { AC } \rightarrow { B },
 { B } \rightarrow { C },
}

Anomalien

Lernkarteikarte: Anomalien

- Update-Anomalie
- Einfüge- oder Insert-Anomalie
- Löschen- oder Delete-Anomalie

Attributhülle

¹Winter, Lindner und Würdinger, *Einführung in relationale Datenbanksysteme & Datenmodellierung*, Seite 191.

Lernkarteikarte: Bestimmung der Attributhülle

```
Hülle(F, α)
  α+ = α
  while(Änderung α+) do
    foreach(Abhängigkeit β → γ ∈ F) do
      if(β ⊆ α+) then α+ = α+ ∪ γ
```

Die Attributhülle α^+ eines bestimmten Attributs α (müsste eigentlich heißen: einer bestimmten Menge von Attributen) ist eine Liste aller Attribute, die von α funktional abhängen.²

Durch die Bestimmung der Attributhülle kann angegeben werden, welche anderen Attribute in einer Relation durch die *gegebenen* Attribute bestimmt werden können.³⁴

gegebenen

Attributhüllenalgorithmus⁵

Beispiel:⁶

FA = {
 { P, Q } → { R },
 { P, R } → { Q },
 { Q } → { S },
 { P, S } → { T },
 { Q, R } → { P },
 { T } → { U },
}

Gesucht

AttrHülle(F, {P,Q})

²Wikipedia-Artikel „Funktionale Abhängigkeit“, Attributhülle.

³Qualifizierungsmaßnahme Informatik - Datenbanksysteme 4, Seite 8.

⁴Winter, Lindner und Würdinger, *Einführung in relationale Datenbanksysteme & Datenmodellierung*, Seite 179.

⁵Winter, Lindner und Würdinger, *Einführung in relationale Datenbanksysteme & Datenmodellierung*, Seite 179.

⁶Qualifizierungsmaßnahme Informatik - Datenbanksysteme 4, Seite 8.

$\{P, Q\}$	Initialisierung von Erg mit den vorgegebenen Attributwerten P und Q
$\{P, Q\} \cup \{R\}$	Die linke Seite der FD $PQ \rightarrow R$ ist bereits in Erg enthalten. Also werden die Attribute der rechten Seite zur Menge Erg hinzugefügt.
$\{P, Q, R\}$	Wegen der FD $PR \rightarrow Q$ käme Q hinzu, ist aber schon in Erg enthalten. Also bleibt Erg unverändert.
$\{P, Q, R\} \cup \{S\}$	Die linke Seite der FD $Q \rightarrow S$ ist bereits in Erg enthalten. Also werden die Attribute der rechten Seite zur Menge Erg hinzugefügt.
$\{P, Q, R, S\} \cup \{T\}$	Die linke Seite der FD $PS \rightarrow T$ ist bereits in Erg enthalten. Also werden die Attribute der rechten Seite zur Menge Erg hinzugefügt.
$\{P, Q, R, S, T\}$	Wegen der FD $QR \rightarrow P$ käme P hinzu, ist aber schon in Erg enthalten. Also bleibt Erg unverändert.
$\{P, Q, R, S, T\} \cup \{U\}$	Die linke Seite der FD $T \rightarrow U$ ist bereits in Erg enthalten. Also werden die Attribute der rechten Seite zur Menge Erg hinzugefügt.
$\{P, Q, R, S, T, U\}$	Es ist also bereits die gesamte Attributmenge erreicht. Weitere Betrachtungen erübrigen sich hiermit.

Die Normalformen

Erste Normalform⁷

Ein Relationenschema R ist in **erster Normalform (1NF)**, wenn alle Attribute *atomar* sind.

atomar

1NF verletzt

FA = {
 $\{CDID\} \rightarrow \{Album, Gründungsjahr, Erscheinungsjahr, Titelliste\},$
 }

in 1NF

FA = {
 $\{CDID, Track\} \rightarrow \{Album, Gründungsjahr, Erscheinungsjahr, Titel\},$
 }⁸

Zweite Normalform

Eine Relation ist genau dann in der zweiten Normalform, wenn die erste Normalform vorliegt und *kein Nichtprimärattribut* (Attribut, das nicht Teil eines Schlüsselkandidaten ist) funktional von einer *echten Teilmenge eines Schlüsselkandidaten abhängt*.

Kein Nichtprimärattribut

echten Teilmenge eines Schlüsselkandidaten abhängt

⁷Winter, Lindner und Würdinger, *Einführung in relationale Datenbanksysteme & Datenmodellierung*, Seite 195.

⁸Wikipedia-Artikel „Normalisierung (Datenbank)“.

Anders gesagt: Jedes nicht-primäre Attribut (nicht Teil eines Schlüssels) ist jeweils von **allen ganzen Schlüsseln** abhängig, nicht nur von einem Teil eines Schlüssels. Wichtig ist hierbei, dass die Nichtschlüsselattribute wirklich von allen Schlüsseln vollständig abhängen.⁹

2NF verletzt

$R = \{$
 $\quad \{ CDID, Track \} \rightarrow \{ Album, Gründungsjahr, Erscheinungsjahr, Titel \},$
 $\quad \{ CDID \} \rightarrow \{ Album, Gründungsjahr, Erscheinungsjahr \},$
 $\}$
 Felder Albumtitel, Interpret und Erscheinungsjahr sind vom Feld CDID abhängig, aber nicht vom Feld Track.

in 2NF

$R_{CD} = \{$
 $\quad \{ CDID \} \rightarrow \{ Albumtitel, Interpret, Gründungsjahr, Erscheinungsjahr \},$
 $\}$
 $R_{Lied} = \{$
 $\quad \{ CDID, Track \} \rightarrow \{ Titel \},$
 $\}$ ¹⁰

Dritte Normalform¹¹

aufgrund von transitiven Abhängigkeiten über Nichtschlüsselkandidaten Falls A nicht prim ist, muss A funktional von jedem Superschlüssel abhängen

Die dritte Normalform ist genau dann erreicht, wenn sich das Relationenschema in der 2NF befindet, und kein Nichtschlüsselattribut von einem Schlüsselkandidaten *transitiv abhängt*.¹²

transitiv abhängt.

3NF verletzt

$R_{CD} = \{$
 $\quad \{ CDID \} \rightarrow \{ Albumtitel, Interpret, Gründungsjahr, Erscheinungsjahr \},$
 $\quad \{ Interpret \} \rightarrow \{ Gründungsjahr \},$
 $\}$
 $R_{Lied} = \{$
 $\quad \{ CDID, Track \} \rightarrow \{ Titel \},$
 $\}$

in 3NF

$R_{CD} = \{$
 $\quad \{ CDID \} \rightarrow \{ Albumtitel, Interpret, Erscheinungsjahr \},$
 $\}$
 $R_{Interpret} = \{$
 $\quad \{ Interpret \} \rightarrow \{ Gründungsjahr \},$
 $\}$

⁹Wikipedia-Artikel „Normalisierung (Datenbank)“, Zweite Normalform (2NF).

¹⁰Wikipedia-Artikel „Normalisierung (Datenbank)“.

¹¹Winter, Lindner und Würdinger, *Einführung in relationale Datenbanksysteme & Datenmodellierung*, Seite 201.

¹²Wikipedia-Artikel „Normalisierung (Datenbank)“, Dritte Normalform (3NF).

$$R_{Lied} = \{ \\ \{ CDID, Track \} \rightarrow \{ Titel \}, \\ \}$$

Boyce-Codd-Normalform

Sei R eine Relationenschema in erster Normalform. Sei F eine Menge einfacher nichttrivialer funktionaler Abhängigkeiten für R . R ist in Boyce-Codd-Normalform (BCNF) bzgl. F , falls für jede FD $X \rightarrow A$ aus F gilt: X ist ein Schlüsselkandidat von R .¹³

Ein Relationenschema ist in der Boyce-Codd-Normalform, wenn es in der 3NF ist und jede Determinante (Attributmenge, von der andere Attribute funktional abhängen) ein Schlüsselkandidat ist (oder die Abhängigkeit ist trivial).

Die BCNF (nach Raymond F. Boyce und Edgar F. Codd) verhindert, dass Teile zweier aus mehreren Feldern zusammengesetzten Schlüsselkandidaten voneinander abhängig sind.

Die Überführung in die BCNF ist zwar immer verlustfrei möglich, aber nicht immer abhängigkeiterhaltend. Die Boyce-Codd-Normalform war ursprünglich als Vereinfachung der 3NF gedacht, führte aber zu einer neuen Normalform, die diese verschärft: Eine Relation ist automatisch frei von transitiven Abhängigkeiten, wenn alle Determinanten Schlüsselkandidaten sind.

BCNF verletzt

$$R_{Sportler} = \{ \\ \{ Name, Verein \} \rightarrow \{ Sportart \}, \\ \{ Name, Sportart \} \rightarrow \{ Verein \}, \\ \{ Verein \} \rightarrow \{ Sportart \}, \\ \}$$

in BCNF

$$R_{Sportler} = \{ \\ \{ Name \} \rightarrow \{ Verein \}, \\ \} \\ R_{Verein} = \{ \\ \{ Verein \} \rightarrow \{ Sportart \}, \\ \}$$

Beispiel

Gegeben ist das Relationenschema *Aufnahmeprüfung*{*PersNr*, *Schuelername*, *Fach*, *Note*} und die Menge von FDs

$$FA = \{ \\ \{ PersNr, Schuelername \} \rightarrow \{ Note \}, \\ \{ Schuelername, Fach \} \rightarrow \{ Note \}, \\ \{ PersNr \} \rightarrow \{ Fach \}, \\ \{ Fach \} \rightarrow \{ PersNr \}, \\ \}.$$

¹³Wikipedia-Artikel „Normalisierung (Datenbank)“, Boyce-Codd-Normalform (BCNF).

Die funktionalen Abhängigkeiten implizieren, dass es zwei Schlüsselkandidaten gibt, nämlich $\{PersNr, Schuelername\}$ und $\{Schuelername, Fach\}$.

Das Relationenschema befindet sich in 3NF, da gilt:

- PersNr, Fach und Schuelername sind prime Attribute.
- Bei allen FDs, auf deren rechten Seite das (einzige) nichtprime Attribut Note vorkommt, ist die Attributmenge auf der linken Seite ein Super-schlüssel.

Trotzdem kann es zu Redundanzen kommen, da die Information, in welchem Fach eine Lehrkraft prüft, mehrfach abgespeichert wird.

Literatur

- [1] Alfons Kemper und André Eickler. *Datenbanksysteme. eine Einführung*. 2013.
- [2] *Qualifizierungsmaßnahme Informatik - Datenbanksysteme 4. Funktionale Abhängigkeiten, Normalformen, Kanonische Überdeckung, Synthesalgorithmus*. https://www.studon.fau.de/file2480907_download.html.
- [3] *Wikipedia-Artikel „Funktionale Abhängigkeit“*. https://de.wikipedia.org/wiki/Funktionale_Abh%C3%A4ngigkeit.
- [4] *Wikipedia-Artikel „Normalisierung (Datenbank)“*. [https://de.wikipedia.org/wiki/Normalisierung_\(Datenbank\)](https://de.wikipedia.org/wiki/Normalisierung_(Datenbank)).
- [5] Stefan Winter, Annabel Lindner und Markus Würdinger. *Einführung in relationale Datenbanksysteme & Datenmodellierung*. https://www.studon.fau.de/file2686598_download.html.