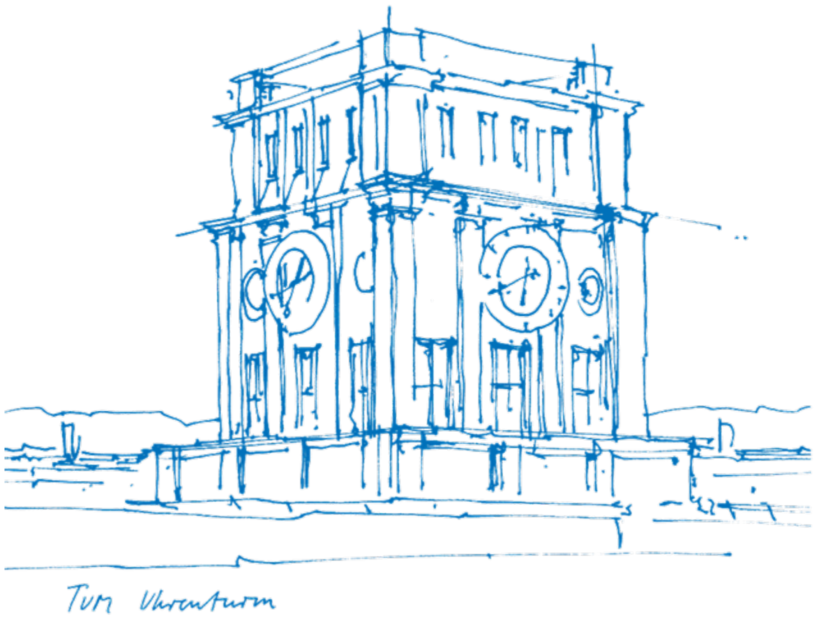
Grundlagen Datenbanken

Benjamin Wagner

7. Dezember 2018



Allgemeines

- Folien von mir sollen unterstützend dienen. Sie sind nicht von der Übungsleitung abgesegnet und haben keinen Anspruch auf Vollständigkeit (oder Richtigkeit).
- Bei Fragen: wagnerbe@in.tum.de
- Vorlesungsbegleitendes Buch von Professor Kemper (Chemiebib)
- Mein Foliensatz ist online: `https://github.com/wagjamin/GDB2018`

Funktionale Abhängigkeiten

- Betrachte Schema $\mathscr{R}$ bestehend aus Relationen $\mathscr{R}_1, \mathscr{R}_2, ..., \mathscr{R}_n$ mit Ausprägung R
- Betrachte **funktionale Abhängigkeit** $\alpha \rightarrow \beta$
- Das heißt: $r, t \in R : r.\alpha = t.\alpha \Rightarrow r.\beta = r.\beta$
- **Frage:** Was bedeutet das in Worten?
- Zu einer Menge funktionaler Abhängigkeiten F kann die Hülle F^+ bestimmt werden

Schlüssel

- Wir erinnern uns: Schlüssel identifizieren Tupel eindeutig
- In der Relation $\mathscr{R}$ ist $\alpha \subseteq \mathscr{R}$ ein **Superschlüssel**, falls: $\alpha \rightarrow \mathscr{R}$
- Volle funktionale Abhängigkeit: α kann nicht weiter verkleinert werden
- Dann heißt α **Kandidatenschlüssel**

Warum machen wir das alles?!

- Wir wollen quantifizieren, ob Schemata gut oder schlecht sind
- Dafür braucht es etwas Theorie
- **Ziel:** Schöne Schemata entwerfen können
- Ab jetzt: Zerlege Relationenschema $\mathscr{R}$ in Schemata $\mathscr{R}_1, \mathscr{R}_2, \ldots, \mathscr{R}_n$
- Invariante: Abhängigkeitserhaltung, Verlustlosigkeit
- Abhängigkeitserhaltung: $F_{\mathscr{R}}^{+} = (F_{\mathscr{R}_1} \cup \ldots \cup F_{\mathscr{R}_n})^{+}$

Verlustlosigkeit

- Eine Zerlegung von $\mathscr{R}$ in $\mathscr{R}_1, \mathscr{R}_2$ heißt verlustlos, wenn mindestens eine der folgenden funktionalen Abhängigkeiten herleitbar ist:
  $\mathscr{R}_1 \cap \mathscr{R}_2 \rightarrow \mathscr{R}_1 \in F_R^+$ oder $\mathscr{R}_1 \cap \mathscr{R}_2 \rightarrow \mathscr{R}_2 \in F_R^+$
- **Beispiel:** Pizaesser

Pizzaesser		
Restaurant	Gast	Pizza
Bella Italia	Ben	Funghi
Pizza Huber	Jonas	Salami
Bella Italia	Jonas	Tonno

- **Frage:** Kann man die Relation verlustlos in {[Restaurant, Gast]} und {[Gast, Pizza]} zerlegen?

Normalformen

- Quantifizieren Qualität der Relation
- **Erste Normalform:** bei uns immer eingehalten: Attribute müssen atomare Werte haben
- **Zweite Normalform:** Eine Relation $\mathscr{R}$ mit FDs F ist in 2NF, falls jedes Nichtschlüssel-Attribut $A \in \mathscr{R}$ von jedem Kandidatenschlüssel in $\mathscr{R}$ voll funktional abhängig ist
- **Dritte Normalform:** Nichtschlüssel-Attribute dürfen nur Fakten von Schlüsseln darstellen
- **Boyce-Codd Normalform:** Informationseinheiten werden nicht mehrmals gespeichert

Dritte Normalform

- Relationenschema $\mathscr{R}$ ist in 3NF, wenn für jede FD $\alpha \to B$ mit $\alpha \subseteq \mathscr{R}$ und $B \in \mathscr{R}$ gilt:
  - $B \in \alpha$, d.h. FD trivial, oder
  - α ist Superschlüssel von $\mathscr{R}$, oder
  - B in Kandidatenschlüssel von $\mathscr{R}$ enthalten
- **Synthesealgorithmus** berechnet verlustlose, abhängigkeitsbewahrende Zerlegung von $\mathscr{R}$ in 3NF
- **Kanonische Überdeckung:** möglichst redundanzfreie Darstellung der FDs einer Relation

Dritte Normalform

Algorithmus 1: Synthesealgorithmus

Data: Relationenschema $\mathscr{R}$, FDs F

Result: Zerlegung $\mathscr{R}_1, \ldots, \mathscr{R}_n$ in 3NF

F_c = kanonische_überdeckung(F) ;

for $(\alpha \rightarrow \beta) \in F_c$ **do**

$\quad \mathscr{R}_\alpha = \alpha \cup \beta$;

$\quad F_a = \{\alpha\prime \rightarrow \beta\prime | \alpha\prime \cup \beta\prime \in \mathscr{R}_\alpha\}$;

end

if *Kein* $\mathscr{R}_\alpha$ *enthält Kandidatenschlüssel* **then**

$\quad \kappa$ = kandidatenschlüssel($\mathscr{R}$) ;

$\quad \mathscr{R}_\kappa = \kappa$;

$\quad F_\kappa = \emptyset$;

end

Teilschemata eliminieren;

Boyce-Codd Normalform

- Relationenschema $\mathscr{R}$ ist in BCNF, wenn für jede FD $\alpha \to \beta$ mit $\alpha, \beta \subseteq \mathscr{R}$ gilt:
  - $\beta \subseteq \alpha$, d.h. FD trivial, oder
  - α ist Superschlüssel von $\mathscr{R}$
- **Dekompositionsalgorithmus** berechnet verlustlose Zerlegung von $\mathscr{R}$ in BCNF
- **Achtung:** es kann nicht garantiert werden, dass die Zerlegung abhängigkeitsbewahrend ist

Boyce-Codd Normalform

Algorithmus 2: Dekompositionsalgorithmus

Data: Relationenschema $\mathscr{R}$, FDs F

Result: Zerlegung $Z = \{\mathscr{R}_1, \ldots, \mathscr{R}_n\}$ in BCNF

$Z = \{\mathscr{R}\}$;

while $\exists \mathscr{R}_i \in Z : \mathscr{R}_i$ *nicht in BCNF* **do**

    repeat

        wähle $(\alpha \to \beta) \in F_{\mathscr{R}_i}$ nicht trivial ;

    until $(\alpha \cap \beta = \emptyset) \wedge !(\alpha \to \mathscr{R}_i)$;

    $\mathscr{R}_{i_1} = \alpha \cup \beta$;

    $\mathscr{R}_{i_2} = \mathscr{R}_i \setminus \beta$;

    $Z = Z \setminus \mathscr{R}_i$;

    $Z = Z \cup \mathscr{R}_{i_1} \cup \mathscr{R}_{i_1}$;

end