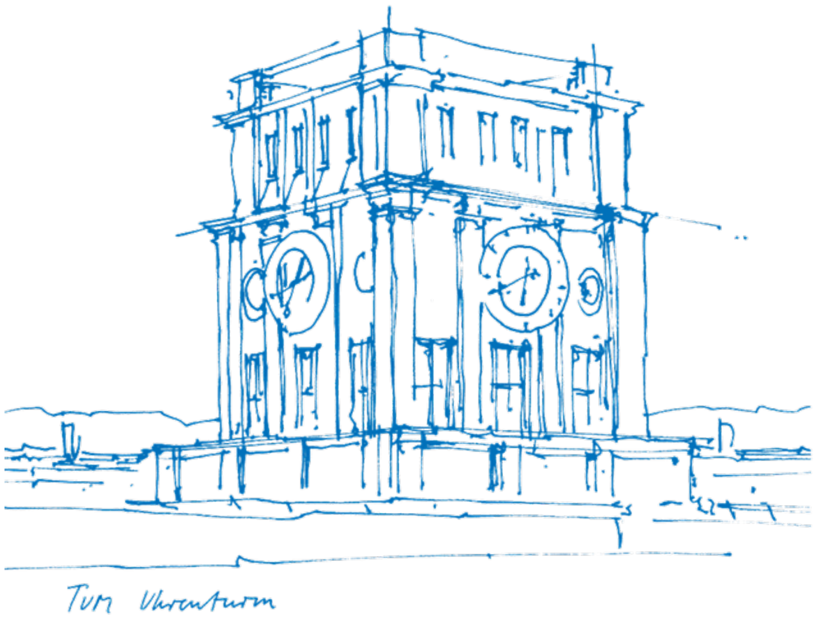
Grundlagen Datenbanken

Benjamin Wagner

13. 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 oder Korrekturvorschlägen: 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?

Attributhülle

- Bestimme maximales $\beta \subseteq \mathscr{R}$, sodass $\alpha \rightarrow \beta$ gilt

Algorithmus 1: Attributhülle

Data: Funktionale Abhägigkeiten F, $\alpha \subseteq \mathscr{R}$

Result: $\beta \subseteq \mathscr{R}$ maximal, sodass $\alpha \rightarrow \beta$

```
abhängig = {α} ;
repeat
    abhängig_alt = abhängig ;
    for β → γ ∈ F do
        if β ⊆ abhängig then
            abhängig = abhängig ∪ γ ;
        end
    end
until abhängig_alt == abhängig;
return abhängig
```

Kanonische Überdeckung

- F_c heißt kanonische Überdeckung von F, wenn gilt:

* $F_c^+ = F^+$
* $(\alpha \to \beta) \in F : \forall A \in \alpha : (F_c \setminus \{\alpha \to \beta\} \cup \{(\alpha \setminus \{A\}) \to \beta\})^+ \neq F_c^+$
* $(\alpha \to \beta) \in F : \forall B \in \beta : (F_c \setminus \{\alpha \to \beta\} \cup \{\alpha \to (\beta \setminus \{B\})\})^+ \neq F_c^+$
* Jede linke Seite einer FD in F_c ist einzigartig

- **Frage:** was bedeuten Bedingung zwei & drei in Worten?

Kanonische Überdeckung

- Linksreduktion macht linke Seiten der FDs so klein wie möglich

Algorithmus 2: Linksreduktion

Data: Funktionale Abhägigkeiten F

Result: Linksreduktion F' von F

$F' = F$;

for $\alpha \to \beta \in F$ **do**

  for $A \in \alpha$ **do**

    if $\beta \subseteq \mathit{Attribut\_Hülle}(F', \alpha \setminus \{A\})$ **then**

      $F' = F' \setminus \{\alpha \to \beta\} \cup \{(\alpha \setminus \{A\}) \to \beta\}$;

    end

  end

end

Kanonische Überdeckung

- Rechtsreduktion macht rechte Seiten der FDs so klein wie möglich

Algorithmus 3: Rechtsreduktion

Data: Funktionale Abhägigkeiten F

Result: Rechtsreduktion F' von F

$F' = F$;

for $\alpha \rightarrow \beta \in F$ **do**

  for $B \in \beta$ **do**

    if $B \in$ *Attribut_Hülle*$(F' \setminus \{\alpha \rightarrow \beta\} \cup \{\alpha \rightarrow (\beta \setminus \{B\})\}, \alpha)$ **then**

      $F' = F' \setminus \{\alpha \rightarrow \beta\} \cup \{\alpha \rightarrow (\beta \setminus \{B\})\}$;

    end

  end

end

Kanonische Überdeckung

- Nun können wir kanonische Überdeckung bestimmen

Algorithmus 4: Kanonische Überdeckung bestimmen

Data: Funktionale Abhägigkeiten F

Result: Kanonische Überdeckung F_c von F

$F_c = F$;

$F_c =$ Linksreduktion(F_c) ;

$F_c =$ Rechtsreduktion(F_c) ;

for $\alpha \rightarrow \emptyset \in F_c$ **do**

  $F_c = F_c \setminus \{\alpha \rightarrow \emptyset\}$;

end

$F_c =$ Gleiche_Linke_Seiten_Zusammenfassen(F_c);

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 5: Synthesealgorithmus

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

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

F_c = kanonische_überdeckung(F) ;

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

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

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

end

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

  κ = kandidatenschlüssel($\mathscr{R}$) ;

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

  $F_\kappa = \emptyset$;

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 6: Dekompositionsalgorithmus BCNF

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

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

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

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

  for $(\alpha \to \beta) \in F_{\mathscr{R}_i}$ *nicht trivial* **do**

    if $(\alpha \cap \beta = \emptyset) \wedge !(\alpha \to \mathscr{R}_i)$ **then**

      break ;

    end

  end

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

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

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

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

end

Mehrwertige Abhängigkeiten

- MVDs verallgemeinern funktionale Abhängigkeiten
- Für $\alpha, \beta \subseteq \mathscr{R}$, schreiben wir: $\alpha \rightarrow\rightarrow \beta$
- Das heißt: für Tupel mit gleichem α kann man β vertauschen und die Tupel bleiben in R
- **Frage:** warum ist das nicht immer erfüllt?

Fähigkeiten

Name	Sprache	ProgSprache
Benjamin	Deutsch	Java
Benjamin	Englisch	C++
Benjamin	Deutsch	C++
Benjamin	Englisch	Java
Kanye	Englisch	LOLCAT

Mehrwertige Abhängigkeiten

- Eine Zerlegung von $\mathscr{R}$ in $\mathscr{R}_1, \mathscr{R}_2$ ist verlustlos, genau dann wenn mindestens eine der folgenden MVDs herleitbar ist:
  $\mathscr{R}_1 \cap \mathscr{R}_2 \rightarrow\rightarrow \mathscr{R}_1$ oder $\mathscr{R}_1 \cap \mathscr{R}_2 \rightarrow\rightarrow \mathscr{R}_2$
- Gibt Regeln, mit denen man aus einer Menge D von MVDs die Hülle D^+ berechnen kann
- MVD $\alpha \rightarrow\rightarrow \beta$ heißt trivial, wenn $\beta \subseteq \alpha$ oder $\beta = \mathscr{R} \setminus \alpha$
- **Frage:** welche MVDs gelten in der Relation zuvor?

Vierte Normalform

- Hier wird zusätzlich zur BCNF noch die Redundanz durch MVDs ausgeschlossen
- Schema $\mathscr{R}$ ist in 4NF, wenn für jede MVD $\alpha \rightarrow\rightarrow \beta \in D^+$ gilt:
  * $\alpha \rightarrow\rightarrow \beta$ ist trivial, oder
  * α ist Superschlüssel von $\mathscr{R}$
- **Dekompositionsalgorithmus** kann wieder verwendet werden, um verlustlose Zerlegung in 4NF zu berechnen
- **Frage:** warum ist eine Relation in 4NF immer auch in BCNF?

Vierte Normalform

Algorithmus 7: Dekompositionsalgorithmus 4NF

Data: Relationenschema $\mathscr{R}$, MVDs D

Result: Zerlegung $Z = \{\mathscr{R}_1, \ldots, \mathscr{R}_n\}$ in 4NF

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

while $\exists \mathscr{R}_i \in Z : \mathscr{R}_i$ *nicht in 4NF* **do**

  for $(\alpha \rightarrow\rightarrow \beta) \in D_{\mathscr{R}_i}$ *nicht trivial* **do**

    if $(\alpha \cap \beta = \emptyset) \wedge !(\alpha \rightarrow\rightarrow \mathscr{R}_i)$ **then**

      break ;

    end

  end

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

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

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

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

end