Übung zu Datenbanken I

Relationaler Entwurf

WS 2019 / 2020

Relationaler Entwurf

- Verfeinerung des logischen Entwurfs durch Beachtung der Integritätsbedingungen (z.B. **funktionale Abhängigkeiten**)
  - Vermeidung von Redundanzen durch Aufspalten von Relationenschemata (**Normalform**), unter der:
    - **Abhängigkeitstreue**: Es werden nur semantisch sinnvolle und konsistente Anwendungsdaten dargestellt. Nachweis über den **RAP-Algorithmus**.
    - **Verbundtreue**: Alle Anwendungsdaten sollen aus Basisrelationen hergeleitet werden können.
- Entwurfsverfahren
  - **Dekomposition**
  - **Synthese**

Funktionale Abhängigkeiten

Definition (Funktionale Abhängigkeit)

Gegeben seien zwei Attributmengen X und Y. Dann existiert eine funktionale Abhängigkeit (FD) zwischen X und Y, wenn der Attributwert x_i unter X den zugehörigen Attributwert y_i unter Y eindeutig bestimmt.

Alternative Schreibweisen:

- $X \rightarrow Y$
- Y ist funktional abhängig von X
- Y hängt funktional von X ab
- X bestimmt Y funktional

Beispiel:

A	B	C
1	1	3
1	1	3
1	2	4

- A, B $\rightarrow$ C, denn $(1, 1) \rightarrow 3$ und $(1, 2) \rightarrow 4$ eindeutig
- A $\nrightarrow$ C, denn $1 \rightarrow 3$ und $1 \rightarrow 4$

Partielle und transitive Abhängigkeit

Definition (Partielle Abhängigkeit)

Gegeben seien zwei Attributmengen X und Y. Dann ist $X \rightarrow Y$ eine partielle Abhängigkeit, wenn gilt:

$$\exists A_i \in X : (X - \{A_i\}) \rightarrow Y$$

Beispiel:
$S = \{(ABC, \{AB\})\}$ mit $F = \{A \rightarrow C\}$

Definition (Transitive Abhängigkeit)

Gegeben seien zwei Attributmengen X und Z. Dann ist $X \rightarrow Z$ eine transitive Abhängigkeit, wenn gilt:

$$\nexists \text{ Attributmenge } Y : X \rightarrow Y, Y \rightarrow Z, Y \nrightarrow X, Z \notin XY$$

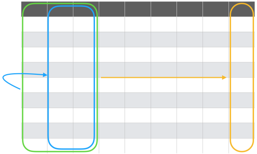
Beispiel:
$S = \{(ABC, \{A\})\}$ mit $F = \{A \rightarrow B, B \rightarrow C\}$

Partielle und transitive Abhängigkeiten für 2NF und 3NF

- Zweite Normalform (2NF):

$$\begin{tikzcd} K \arrow[r, bend left] & Y \arrow[l, bend left, "\times" description, "Y \subseteq K"'] \arrow[r] & A \notin KY \end{tikzcd}$$

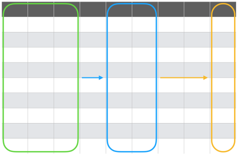
Schlüssel Nichtprimattribut



- Dritte Normalform (3NF):

$$\begin{tikzcd} K \arrow[r, bend left] & Y \arrow[l, bend left, "\times" description] \arrow[r] & A \notin KY \end{tikzcd}$$

Schlüssel Nichtprimattribut



Dritte Normalform — **S1**

Definition (Erste Normalform)

- Erlaubt im Relationenschema nur atomare Attribute
  $\rightarrow$ Attributwerte sind Elemente von Standardtypen wie **string** oder **integer**, keine Konstruktoren wie **array** oder **set**
- Eliminierung mengenwertiger Attribute durch Duplizierung

Definition (Zweite Normalform)

Erlaubt keine partiellen Abhängigkeiten eines Nicht-Primattributs von einem Schlüssel

Definition (Dritte Normalform)

Erlaubt keine transitiven Abhängigkeiten eines Nicht-Primattributs von einem Schlüssel

Normalformen — Beispiele

Aufgabe:

- $S_1 = \{(ABC, \{A\})\}$ mit $F_1 = \{A \to B, A \to C\}$
- $S_2 = \{(ABC, \{A\})\}$ mit $F_2 = \{A \to B, B \to C\}$
- $S_3 = \{(ABC, \{AB\})\}$ mit $F_3 = \{A \to B, B \to C\}$
- $S_4 = \{(AB, \{A\}), (BC, \{B\})\}$ mit $F_4 = \{A \to B, B \to C\}$
- $S_5 = \{(ABC, \{A, B\})\}$ mit $F_5 = \{A \to B, B \to C, B \to A\}$

Normalformen — Beispiele

Aufgabe:

- $S_1 = \{(ABC, \{A\})\}$ mit $F_1 = \{A \rightarrow B, A \rightarrow C\}$
- $S_2 = \{(ABC, \{A\})\}$ mit $F_2 = \{A \rightarrow B, B \rightarrow C\}$
- $S_3 = \{(ABC, \{AB\})\}$ mit $F_3 = \{A \rightarrow B, B \rightarrow C\}$
- $S_4 = \{(AB, \{A\}), (BC, \{B\})\}$ mit $F_4 = \{A \rightarrow B, B \rightarrow C\}$
- $S_5 = \{(ABC, \{A, B\})\}$ mit $F_5 = \{A \rightarrow B, B \rightarrow C, B \rightarrow A\}$

Lösung:

	S_1	S_2	S_3	S_4	S_5
1NF	✓	✓	✓	✓	✓
2NF	✓	✓	–	✓	✓
3NF	✓	–	–	✓	✓

Minimalität — **S2**

Definition (Minimalität)

Minimale Anzahl von Relationenschemata, die die anderen Eigenschaften (**S1**, **T1**, **T2**) erfüllt

Aufgabe:

- $F = \{A \to BC, C \to D, C \to A, E \to C\}$
- $S = \{(ABCD, \{A, C\}), (CE, \{E\})\}$

Minimalität — S2

Definition (Minimalität)

Minimale Anzahl von Relationenschemata, die die anderen Eigenschaften (**S1**, **T1**, **T2**) erfüllt

Aufgabe:

- $F = \{A \to BC, C \to D, C \to A, E \to C\}$
- $S = \{(ABCD, \{A, C\}), (CE, \{E\})\}$

Lösung:

Wähle ein kleineres Schema:

- $S' = \{(ABCDE, \{E\})\}$

$\Rightarrow$ S' enthält transitive Abhängigkeit $E \to C \to A$

$\Rightarrow$ S' erfüllt **S1** nicht mehr

S ist minimal

Membership Problem

Kann eine bestimmte FD $X \to Y$ aus der vorgegebenen Menge F abgeleitet werden, d.h. wird sie von F impliziert?

- Hülle: $F_R^+ := \{f \mid (f \text{ FD über } R) \wedge F \models f\}$

Membership-Problem (1): $X \to Y \in F_R^+$?

- Hülle von X bzgl. F ist $X_F^+ := \{A \mid X \to A \in F^+\}$

Membership-Problem (2): $Y \subseteq X_F^+$?

RAP-Algorihtmus

- RAP-Regeln:

R	Reflexivität	$\{\} \Rightarrow X \rightarrow X$
A	Akkumulation	$\{X \rightarrow YZ, Z \rightarrow AW\} \Rightarrow X \rightarrow YZA$
A'		$\{X \rightarrow YZ, Z \rightarrow W\} \Rightarrow X \rightarrow YZW$
P	Projektivität	$\{X \rightarrow YZ\} \Rightarrow X \rightarrow Y$

- Die RAP-Regelmenge ist vollständig, die Regeln selbst sind gültig und unabhängig.
  - gültig (sound): Regeln leiten keine FDs ab, die logisch nicht impliziert werden
  - vollständig (complete): alle implizierten FDs werden abgeleitet
  - unabhängig (independent): keine Regel kann weggelassen werden

RAP-Algorithmus

Algorithm 1 RAP-Algorihtmus

$X^+ := X$ ▷ R-Regel
repeat
  $X_{\text{Hilfe}} := X^+$
  for all FDs $Y \to Z \in F$ **do**
    if $X_i \to Y_i \in F$ mit $X_i \subseteq X^+$ **then** $X^+ := X^+ \cup Y_i$
    end if ▷ A'-Regel
  end for
until $X^+ = X_{\text{Hilfe}}$
if $Y \subseteq X^+$ **then** $X \to Y \in F^+$ ▷ P-Regel
end if

- R-Regel $\Rightarrow$ Ausnutzten der Reflexivität
- A-Regel (mehrfach angewandt) $\Rightarrow$ Berechnung der Hülle
- P-Regel $\Rightarrow$ Ableiten möglicher FDs

Abhängigkeitstreue — **T1**

Definition (Abhängigkeitstreue)

Eine Menge von Abhängigkeiten kann äquivalent in eine zweite Menge von Abhängigkeiten transformiert werden.

Spezieller: Für ein gegebenes Datenbankschema S ist die Menge der (funktionalen) Abhängigkeiten F äquivalent zur Menge der Schlüsselbedingungen G.

Zu zeigen:

- $X \rightarrow Y \in F \Rightarrow X \rightarrow Y \in G^+$
- $X \rightarrow Y \in G \Rightarrow X \rightarrow Y \in F^+$

Abhängigkeitstreue — Beispiel

Aufgabe:

- Datenbankschema: $S = \{(OSHP, \{OSH\})\}$
- Funktionale Abhängigkeiten:

$$F = \{OSH \to P, P \to O\}$$

- Schlüsselabhängigkeiten: $G = \{OSH \to OSHP\}$

$\Rightarrow$ Frage: Ist F äquivalent zu G?

Abhängigkeitstreue — Beispiel

Aufgabe:

- Datenbankschema: $S = \{(OSHP, \{OSH\})\}$
- Funktionale Abhängigkeiten:

$$F = \{OSH \to P, P \to O\}$$

- Schlüsselabhängigkeiten: $G = \{OSH \to OSHP\}$

$\Rightarrow$ Frage: Ist F äquivalent zu G?

Lösung:

- $OSH \to OSHP \in F^+$
- $OSH \to P \in G^+$
- $P \to O \notin G^+$

$\Rightarrow$ Antwort: F ist nicht äquivalent zu G.

Verbundtreue — **T2**

Definition (Verbundtreue)

Für eine Zerlegung eines Relationenschemas R in $\bigcup_{i=1}^{n} R_i$ gilt:

$$R = R_1 \bowtie R_2 \bowtie \ldots \bowtie R_n$$

Kriterium 1 (für 2 Relationenschemata):

- Die Dekomposition einer Attributmenge X in X_1 und X_2 ist verbundtreu bzgl. F über X, wenn

$$X_1 \cap X_2 \rightarrow X_1 \in F^+ \text{ oder } X_1 \cap X_2 \rightarrow X_2 \in F^+$$

Kriterium 2 (allgemeiner, falls Abhängigkeitstreue vorhanden ist):

- Der Universalschlüssel muss in einem Relationenschema der Dekomposition vollständig enthalten sein.
- Universalschlüssel: minimale Teilmenge von U, die U funktional bestimmt

Verbundtreue — Beispiele

Kriterium 1:

- $S_1 = \{(ABC, \{A\})\}$
- $S_2 = \{(AB, \{A, B\}), (BCD, \{C\})\}$
- $S_3 = \{(AB, \{A\}), (BC, \{C\})\}$
- $S_4 = \{(AB, \{A\}), (CD, \{C\})\}$
- $S_5 = \{(ABD, \{B\}), (ABC, \{C\})\}$

Kriterium 2 (Schlüsselsuche):

- $F_1 = \{A \to B, B \to C, A \to D, D \to H, B \to E\}$,
  $U_1 = \{ABCDEH\}$
- $F_2 = \{A \to BCD, BC \to A, BC \to D\}$, $U_2 = \{ABCDE\}$

Verbundtreue — Beispiele

Kriterium 1:

- $S_1 = \{(ABC, \{A\})\}$
- $S_2 = \{(AB, \{A, B\}), (BCD, \{C\})\}$
- $S_3 = \{(AB, \{A\}), (BC, \{C\})\}$
- $S_4 = \{(AB, \{A\}), (CD, \{C\})\}$
- $S_5 = \{(ABD, \{B\}), (ABC, \{C\})\}$

Schema	Verbund-treue
S_1	✓
S_2	✓
S_3	x
S_4	x
S_5	✓

Kriterium 2 (Schlüsselsuche):

- $F_1 = \{A \rightarrow B, B \rightarrow C, A \rightarrow D, D \rightarrow H, B \rightarrow E\}$,
  $U_1 = \{ABCDEH\}$
  $\Rightarrow$ Universalschlüssel: A
- $F_2 = \{A \rightarrow BCD, BC \rightarrow A, BC \rightarrow D\}$, $U_2 = \{ABCDE\}$
  $\Rightarrow$ Universalschlüssel: AE

Beispiel I — Eigenschaften

Aufgabe:

- $F = \{LT \to A, L \to O, O \to E\}$
- $S = \{(LTA, \{LT\}), (LO, \{L\}), (OE, \{O\})\}$

Beispiel I — Eigenschaften

Aufgabe:

- $F = \{LT \to A, L \to O, O \to E\}$
- $S = \{(LTA, \{LT\}), (LO, \{L\}), (OE, \{O\})\}$

Eigenschaften:

- Abhängigkeitstreue: ✓
- 3 Normalform: ✓
  - partielle Abhängigkeit: x
  - transitive Abhängigkeit: x
- Verbundtreue: ✓
  - Der Universalschlüssel LT ist Teil eines Relationenschemas.
- Minimalität: ✓

Beispiel II — Eigenschaften

Aufgabe:

- $F = \{A \to BC, C \to AD, E \to C\}$
- $S = \{(ABCD, \{A\}), (AE, \{E\})\}$

Beispiel II — Eigenschaften

Aufgabe:

- $F = \{A \to BC, C \to AD, E \to C\}$
- $S = \{(ABCD, \{A\}), (AE, \{E\})\}$

Eigenschaften:

- Abhängigkeitstreue: x
  - $C \to AD \notin G^+$
- 3 Normalform: ✓
  - partielle Abhängigkeit: x
  - transitive Abhängigkeit: x, C ist alternativer Schlüssel
- Verbundtreue:
  - ✓ nach Kriterium 1
  - Kriterium 2 hier nicht anwendbar.
- Minimalität: x
  - Setzt Abhängigkeitstreue voraus.

Entwurfsverfahren

Ziel: Universum U und FD-Menge $F \Rightarrow$ lokal erweitertes Datenbankschema $S = \{(R_1, K_1), ..., (R_p, K_p)\}$

Bedingungen:

T1 S charakterisiert F vollständig.
S1 S ist bzgl. F in 3NF.
T2 Die Dekomposition von U in $R_1, ..., R_p$ ist bzgl. F verbundtreu.
S2 Minimalität, d.h. $\nexists S' : S'$ erfüllt **T1**, **S1**, **T2** und $\mid S' \mid < \mid S \mid$.

Verfahren:

- Dekomposition
- Synthese

Dekomposition

- Verfahren:
  - Gegeben: Universalrelationenschema $R = (U, K(F))$ mit $K(F) = \{K \to U : K \to U \in F^+ \text{ und } K \text{ minimal}\}$
  - Gesucht: Zerlegung in $D = \{R_1, R_2, ...\}$ von 3NF-Schemata
- Reihenfolgeabhängig
- Komplexität: NP-vollständiges Problem (Schlüsselsuche in Originalrelation mit exponentiellem Aufwand)

Kennung	Eigenschaft	erfüllt
S1	3NF	$\checkmark$
S2	Minimalität	x
T1	Abhängigkeitstreue	x
T2	Verbundtreue	$\checkmark$

Beispiel — Dekomposition

Aufgabe:

- $F = \{A \to B, B \to C, A \to D, D \to H, B \to E\}$
  $U = ABCDEH$

Beispiel — Dekomposition

Aufgabe:

- $F = \{A \to B, B \to C, A \to D, D \to H, B \to E\}$
  $U = ABCDEH$

Lösung:

- Universalschlüssel: A, denn $A^+ = ABCDEH$
- Universalrelationenschema: $S = \{(ABCDEH, \{A\})\}$
- Transitive Abhängigkeiten von A:
  - $A \to B \to C \Rightarrow \{(ABDEH, \{A\}), (BC, \{B\})\}$
  - $A \to B \to E \Rightarrow \{(ABDH, \{A\}), (BC, \{B\}), (BE, \{B\})\}$
  - $A \to D \to H \Rightarrow \{(ABD, \{A\}), (BC, \{B\}), (BE, \{B\}), (DH, \{D\})\}$
- Ergebnisrelationenschema:
  $S' = \{(ABD, \{A\}), \underbrace{(BC, \{B\}), (BE, \{B\})}_{\text{Minimalität für } (BCE, \{B\})}, (DH, \{D\})\}$

Synthese

- Verfahren:
  - Gegeben: Relationenschema R mit FDs F
  - Gesucht: abhängigkeitstreue, minimale Zerlegung $D = \{R_1, R_2, \ldots\}$ von 3NF-Schemata
- Reihenfolgeunabhängig
- Komplexität: quadratisch

Kennung	Eigenschaft	erfüllt
S1	3NF	✓
S2	Minimalität	✓
T1	Abhängigkeitstreue	✓
T2	Verbundtreue	✓

Zusammenfassung

Kennung	Eigenschaft	Kurzcharakteristik
S1	3NF	keine transitiven Abhängigkeiten eines Nicht-Primattributes von einem Schlüssel
S2	Minimalität	minimale Anzahl von Relationenschemata, die die anderen Eigenschaften erfüllt
T1	Abhängigkeitstreue	alle gegebenen Abhängigkeiten sind durch Schlüssel repräsentiert
T2	Verbundtreue	Originalrelationen können durch den Verbund der Basisrelationen wiedergewonnen werden

Entwurfsverfahren:

- Dekomposition
- Synthese