Datenmodellierung/Datenbanksysteme

VU 184.685/VU 184.686, WS 2022

Relationale Entwurfstheorie – Funktionale Abhängigkeiten

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

Ziele

- Finetuning des relationalen Schemas
- Qualität eines Relationenschemas:
  - Einhaltung von Konsistenzbedingungen
  - Vermeidung von Redundanzen
- Modellierung mittels Datenabhängigkeiten
  - Funktionale Abhängigkeiten
  - Inklusionsabhängigkeiten
  - Verbundabhängigkeiten

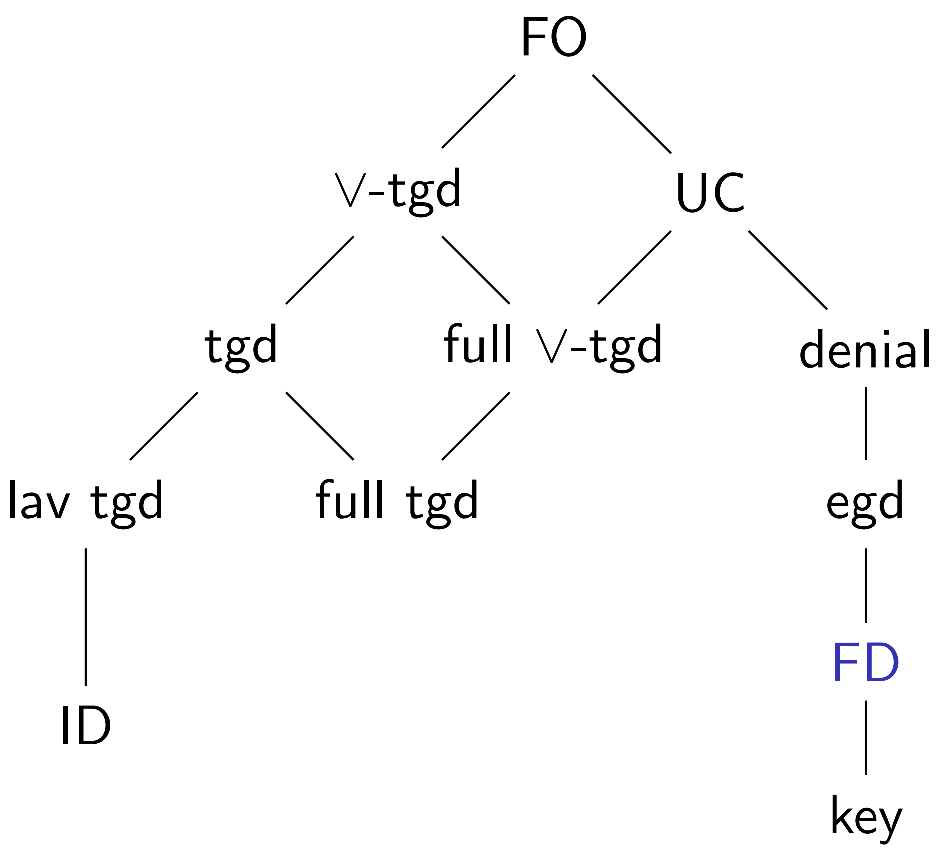
Ziele

- Grundlage: Funktionale Abhängigkeiten (FDs)
  - Motivation
  - Definition
  - Bestimmung
  - Hülle
  - Kanonische Überdeckung
  - Schlüssel
- Normalformen als Gütekriterium
- Ggfls. Verbesserung eines Relationenschemas
  - Synthesealgorithmus
  - Dekomposition

Funktionale Abhängigkeiten

- Motivation
- Definition
- Bestimmung
- Hülle
- Äquivalenz
- Kanonische Überdeckung
- Schlüssel

Integritätsbedingungen: Ein essenzielles DB-Werkzeug



Einsatzgebiete Funktionaler Abhängigkeiten

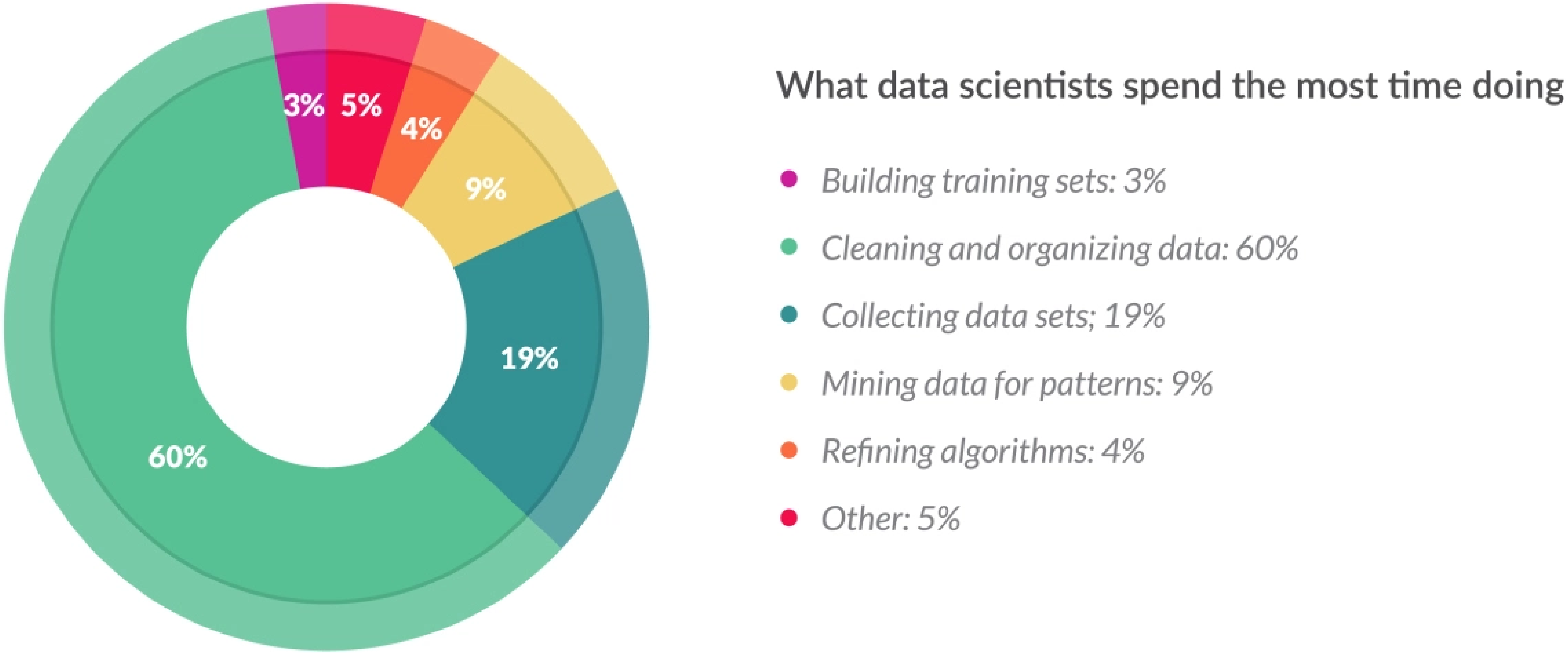
Vorteile Funktionaler Abhängigkeiten:

- Einfache Semantik
- Viele Probleme sind effizient berechenbar

Einsatz von Funktionalen Abhängigkeiten für/in:

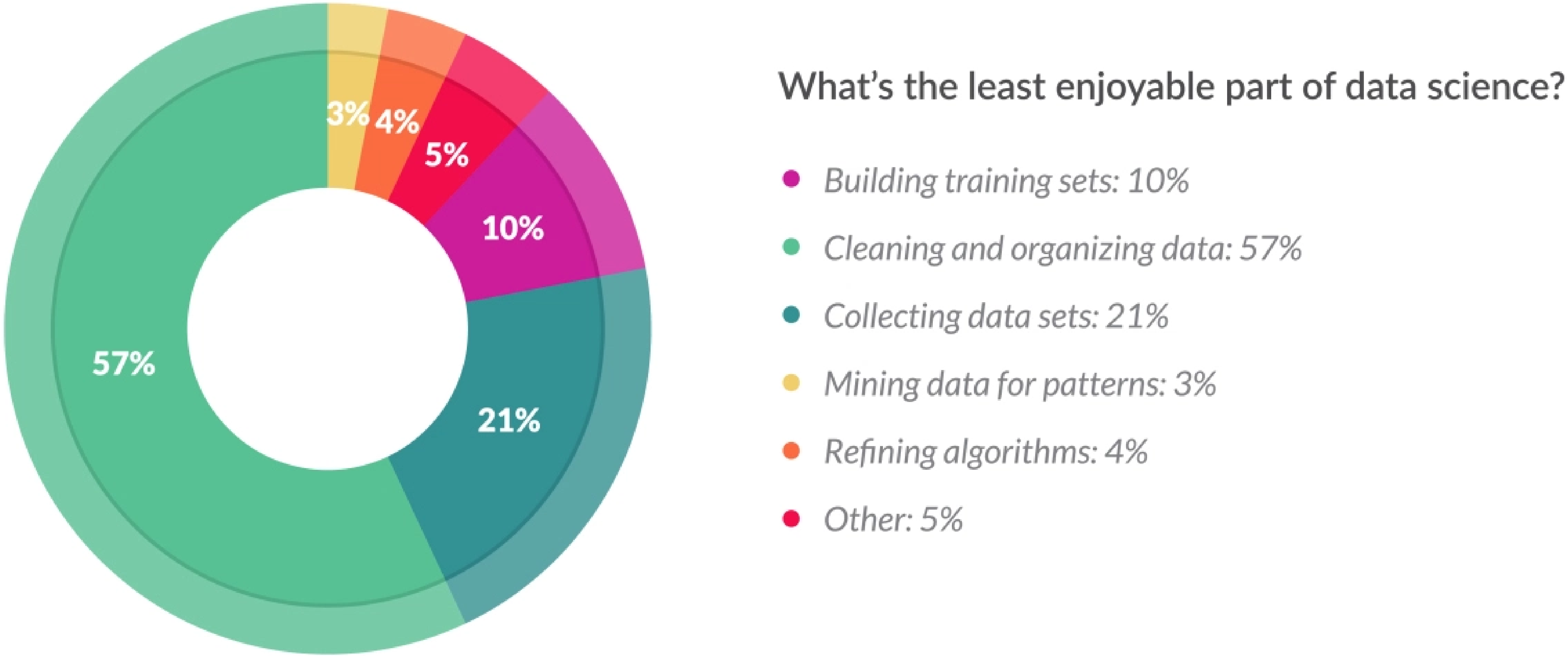
- Konsistenzbedingung für Datenbanken
- Bestimmung der Qualität von Relationenschemata
- "Data Exchange" und "Data Integration"
- "Data Cleaning"
- ...

Anwendungsbeispiel: Data Cleaning



(Forbes, Mai 2016)
(Danke an Emanuel Sallinger)

Anwendungsbeispiel: Data Cleaning



(Forbes, Mai 2016)
(Danke an Emanuel Sallinger)

Inhalt

Funktionale Abhängigkeiten – Notation

Notation:

- Schema $\mathcal{R} = \{A, B, C, D, \ldots, H\}$
- Attribute: $A, B, C, \ldots$ Attributmengen: $\alpha, \beta, \ldots$
- Relation: R, Tupel: $r, s, t, \ldots$ Projektion: $r.\alpha, t.\beta, \ldots$

- Mengendifferenz (für $a \in \mathcal{A}$): $\mathcal{A} - a$ statt $\mathcal{A} \setminus \{a\}$

Funktionale Abhängigkeiten

Definition (Funktionale Abhängigkeit)

Sei $\mathcal{R}$ ein Relationenschema und $\alpha \subseteq \mathcal{R}, \beta \subseteq \mathcal{R}$.
Eine Funktionale Abhängigkeit (FD) ist eine Beziehung $\alpha \to \beta$.

"α bestimmt β"

Definition (Semantik von Funktionalen Abhängigkeiten)

Eine Relation R erfüllt eine funktionale Abhängigkeit (FD) $\alpha \to \beta$ genau dann, wenn für alle Tupel $r, t \in R$ mit $r.\alpha = t.\alpha$ gilt: $r.\beta = t.\beta$.

$\alpha \to \beta$: wenn zwei Tupel die gleichen Werte für alle Attribute in α haben, so haben sie die gleichen Werte auch für alle Attribute in β.

"Die α-Werte bestimmen die β-Werte eindeutig (funktional)"

Funktionale Abhängigkeiten

Beispiel

- Ein Familienstammbaum impliziert folgende FDs:
  - Kind $\rightarrow$ Vater, Mutter
  - Kind, Oma $\rightarrow$ Opa
  - Kind, Opa $\rightarrow$ Oma

Stammbaum

Kind	Mutter	Vater	Oma	Opa
Sofie	Sabine	Alfons	Linde	Lothar
Sofie	Sabine	Alfons	Lisa	Hubert
Niklas	Sabine	Alfons	Linde	Lothar
Niklas	Sabine	Alfons	Lisa	Hubert
Sofie	...	...	Linde	Willi

Funktionale Abhängigkeiten

Beispiel

- Schema $\mathcal{R} = \{A, B, C, D\}$ der Relation R
- Frage: sind folgende FDs auf der Relation R gültig oder nicht:

$\{A\} \rightarrow \{B\}$	ja
$\{C, D\} \rightarrow \{B\}$	ja
$\{B\} \rightarrow \{C\}$	nein
$\{A, B\} \rightarrow \{C\}$	ja
$\{B, C\} \rightarrow \{A\}$	nein
$\{B\} \rightarrow \{A\}$	nein

R			
A	*B*	*C*	*D*
a4	b2	c4	d3
a1	b1	c1	d1
a1	b1	c1	d2
a2	b2	c3	d2
a3	b2	c4	d3

Einhaltung einer FD

Gesucht: Vorgehen zur Überprüfung, ob eine gegebene Relation R die FD $\alpha \to \beta$ erfüllt:

- Die folgende SQL-Abfrage liefert ein leeres Ergebnis:

```
select * from R r1, R r2
where r1.α = r2.α and r1.β ≠ r2.β;
```

```
select * from R r1 where exists(
  select * from R r2 where r1.α = r2.α
                       and r1.β ≠ r2.β);
```

- Für alle möglichen Werte c enhält das Ergebnis der Abfrage

$$\pi_\beta(\sigma_{\alpha=c}(R))$$

maximal ein Tupel

Einhaltung einer FD

Algorithmus zur Überprüfung, ob eine gegebene
Relation R die FD $\alpha \to \beta$ erfüllt:

Input: $(R, \alpha \to \beta)$: Relation R und eine FD $\alpha \to \beta$

Output: ja, wenn FD erfüllt, nein sonst

Einhaltung $(R, \alpha \to \beta)$

- sortiere R nach den Werten von α
- falls alle Gruppen von Tupeln mit gleichen α Werten die gleichen β Werte aufweisen: output(ja) sonst output(nein)

Laufzeit von Einhaltung ist bestimmt durch Aufwand für die Sortierung - $O(nlogn)$

Inhalt

Bestimmung von FDs

Beispiel

Gegeben: Information über ProfessorInnen anhand folgender Attribute:
ProfessorInnen:{[PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, BLand, EW, Landesregierung]}
Frage: Welche funktionalen Abhängigkeiten lassen sich aufgrund der Semantik der zu modellierenden Miniwelt finden?

- PersNr ist ein Kandidatenschlüssel:
  {PersNr} $\rightarrow$ {PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, BLand, EW, Landesregierung}
- Ortsnamen sind innerhalb eines Bundeslandes eindeutig:
  {Ort, BLand} $\rightarrow$ {EW, Vorwahl}

Bestimmung von FDs

Beispiel

- Die Postleitzahl identifiziert einen Ort, das Bundesland und die Einwohnerzahl:
  $\{\text{PLZ}\} \to \{\text{BLand, Ort, EW}\}$
- Die Postleitzahl ändert sich innerhalb der Straße eines Ortes nicht:
  $\{\text{BLand, Ort, Straße}\} \to \{\text{PLZ}\}$
- Landesregierung speichert die Partei des Landeshauptmanns/frau:
  $\{\text{BLand}\} \to \{\text{Landesregierung}\}$
- In einem Raum kann nur einE ProfessorIn sitzen:
  $\{\text{Raum}\} \to \{\text{PersNr}\}$

Inhalt

Hülle einer Menge von Attributen bzw. FDs

Gegeben: Menge F von FDs.

Beispiel: Raum $\rightarrow$ PersNr, PersNr $\rightarrow$ Name

Frage 1: Gegeben zusätzlich eine Menge γ an Attributen, welche Attribute sind von γ durch F funktional bestimmt?

Frage 2: Welche weiteren FDs können aus F abgeleitet werden?

Hülle einer Attributmenge

Gegeben: Eine Menge γ von Attributen und F eine Menge von FDs.

Frage: Welche Attribute sind von γ durch F funktional bestimmt?

Beispiel: $\{\text{Raum} \rightarrow \text{PersNr}, \text{PersNr} \rightarrow \text{Name}\}, \{\text{Raum}\}$
$\Rightarrow \{\text{Raum}, \text{PersNr}, \text{Name}\}$

Definition (Hülle einer Attributmenge)

Die Menge γ^+ der Attribute welche von γ funktional abhängen nennt man die Hülle der Attributmenge γ.

Hülle einer Attributmenge

Berechnung mittels Algorithmus AttrHülle:

Input: (F, γ): Menge F von FDs und Menge γ von Attributen

Output: Menge der Attribute γ^+.

AttrHülle (F, γ)

$\gamma^+ = \gamma$

while $\exists(\alpha \to \beta) \in F$ mit $\alpha \subseteq \gamma^+$ und $\beta \not\subseteq \gamma^+$ **do**

$\quad \gamma^+ := \gamma^+ \cup \beta$

return(γ^+)

Hülle einer Attributmenge

Beispiel

Sei $F = \{RS \rightarrow T, U \rightarrow VX, RX \rightarrow W, T \rightarrow RU\}$

AttrHülle$(F, \{T\})$: $\{R, T, U, V, W, X\}$

AttrHülle$(F, \{RS\})$: $\{R, S, T, U, V, W, X\}$

Hülle von FDs

Gegeben: Menge F von FDs.

Beispiel: Raum $\rightarrow$ PersNr, PersNr $\rightarrow$ Name

Frage 1: Gegeben zusätzlich eine Menge γ an Attributen, welche Attribute sind von γ durch F funktional bestimmt?

Frage 2: Welche weiteren FDs können aus F abgeleitet werden?

Hülle von FDs

Problem: F Menge von FDs. Welche weiteren FDs können daraus abgeleitet werden?

Beispiel: Raum $\rightarrow$ PersNr, PersNr $\rightarrow$ Name $\Rightarrow$ Raum $\rightarrow$ Name

Definition ($F_1 \models F_2$)

Die Menge F_2 von FDs ist aus der Menge F_1 von FDs ableitbar, wenn jede Relation R welche alle FDs in F_1 erfüllt auch alle FDs in F_2 erfüllt.

Definition (Hülle von FDs)

Die Menge aller aus F ableitbaren FDs wird Hülle F^+ von F genannt.

Verständnishilfe aus der Mathematik: V eine Menge von Vektoren. Die Menge aller Vektoren, die aus V mittels Linearkombinationen erhalten werden: lineare Hülle von V.

Ableiten von FDs mittels Attributhülle

FD $\alpha \to \beta$: Die Werte für α bestimmen die Werte für β funktional.

Attributhülle: Alle Attribute γ^+, deren Werte von γ funktional durch F bestimmt sind.

Theorem

Gegeben eine Menge F an FDs und eine Menge γ an Attributen. Dann gilt:

$$F \models \{\gamma \to \mathit{AttrHülle}(F, \gamma)\}$$

Außerdem: $F \models \{\alpha \to \beta\} \Leftrightarrow \beta \subseteq \text{AttrHülle}(F, \alpha)$

Ableiten von FDs mit Hilfe der Armstrong Axiome

Berechnung der Hülle F^+ von F mittels der Armstrong Axiome (1974).

Theorem

Die Armstrong Axiome sind vollständig (erzeugen alle implizierten FDs) und korrekt (erzeugen nur gültige FDs)

Die Armstrong Axiome

Reflexivität: Sei β eine Teilmenge von α ($\beta \subseteq \alpha$) dann gilt immer $\alpha \to \beta$. Insbesondere gilt $\alpha \to \alpha$.

Verstärkung: Falls $\alpha \to \beta$ gilt, dann gilt auch $\alpha\gamma \to \beta\gamma$.

Transitivität: Falls $\alpha \to \beta$, $\beta \to \gamma$ gelten, dann gilt auch $\alpha \to \gamma$.

Zusätzliche Axiome, die die Herleitung der Hülle erleichtern:

Vereinigung: $\alpha \to \beta$ und $\alpha \to \gamma \Rightarrow \alpha \to \beta\gamma$.

Dekomposition: $\alpha \to \beta\gamma \Rightarrow \alpha \to \beta$ und $\alpha \to \gamma$

wichtig: kann immer erreichen, dass rechts nur ein Attribut steht.

Pseudotransitivität: $\alpha \to \beta$ und $\gamma\beta \to \delta \Rightarrow \alpha\gamma \to \delta$.

Die Armstrong Axiome

Beispiel

Herleitung der FD $\{\text{PLZ}\} \rightarrow \{\text{Landesregierung}\}$ aus den restlichen FDs im Beispielschema Professoren:

Es gelten: $\{\text{PLZ}\} \rightarrow \{\text{BLand, Ort, EW}\}$ und
$\{\text{BLand}\} \rightarrow \{\text{Landesregierung}\}$

Dekomposition von $\{\text{PLZ}\} \rightarrow \{\text{BLand, Ort, EW}\}$:

$$\{\text{PLZ}\} \rightarrow \{\text{BLand}\},\ \{\text{PLZ}\} \rightarrow \{\text{Ort}\},\ \{\text{PLZ}\} \rightarrow \{\text{EW}\}$$

Transitivität von $\{\text{PLZ}\} \rightarrow \{\text{BLand}\}$, $\{\text{BLand}\} \rightarrow \{\text{Landesregierung}\}$:

$$\{\text{PLZ}\} \rightarrow \{\text{Landesregierung}\}$$

Die Armstrong Axiome

Beispiel

Herleitung der Vereinigung aus Reflexivität, Verstärkung und Transitivität:

Gegeben: $\alpha \rightarrow \beta$, $\alpha \rightarrow \gamma$

Zu Zeigen: $\alpha \rightarrow \beta\gamma$

Schritt 1: Verstärkung von $\alpha \rightarrow \beta$: $\alpha \rightarrow \alpha\beta$

Schritt 2: Verstärkung von $\alpha \rightarrow \gamma$: $\alpha\beta \rightarrow \beta\gamma$

Schritt 3: Transitivität von $\alpha \rightarrow \alpha\beta$, $\alpha\beta \rightarrow \beta\gamma$: $\alpha \rightarrow \beta\gamma$

Inhalt

- Motivation
- Definition
- Bestimmung
- Hülle
- Äquivalenz
- Kanonische Überdeckung
- Schlüssel

Äquivalenz von Mengen von FDs

Gegeben: Mengen F_1, F_2 von FDs.

Frage: Beschreiben F_1 und F_2 die selbe Menge an FDs?

Äquivalenz von Mengen von FDs

Definition (Äquivalenz von FDs)

Zwei Mengen F, G von FDs sind äquivalent ($F \equiv G$), wenn sie dieselbe Hülle besitzen, d.h. $F^+ = G^+$.

Mathematik: zwei Mengen von Vektoren sind "äquivalent", wenn sie denselben Vektorraum aufspannen

Natürlich gilt: $F \equiv G$ genau dann wenn

- $F \subseteq G^+$ und
- $G \subseteq F^+$.

Äquivalenz von Mengen von FDs

Beispiel (Armstrong Axiome)

$F_1 = \{A \to B, B \to C, C \to A\}$

$F_2 = \{B \to A, C \to B, A \to C\}$

$F_1 \equiv F_2?$

Gilt $F_1 \subseteq F_2^+$?

$A \to B$: ✓ Transitivität von $A \to C$, $C \to B$

$B \to C$: ✓ Transitivität von $B \to A$, $A \to C$

$C \to A$: ✓ Transitivität von $C \to B$, $B \to A$

Gilt $F_2 \subseteq F_1^+$?

$B \to A$: ✓ Transitivität von $B \to C$, $C \to A$

$C \to B$: ✓ Transitivität von $C \to A$, $A \to B$

$A \to C$: ✓ Transitivität von $A \to B$, $B \to C$

Äquivalenz von Mengen von FDs

Beispiel (Attributhülle)

$F_1 = \{A \to B, B \to C, C \to A\}$ $\qquad$ $F_2 = \{B \to A, C \to B, A \to C\}$

$$F_1 \equiv F_2?$$

Gilt $F_1 \subseteq F_2^+$?

$A \to B$: ✓
$B \in \text{AttrHülle}(F_2, \{A\})$

$B \to C$: ✓
$C \in \text{AttrHülle}(F_2, \{B\})$

$C \to A$: ✓
$A \in \text{AttrHülle}(F_2, \{C\})$

Gilt $F_2 \subseteq F_1^+$?

$B \to A$: ✓
$A \in \text{AttrHülle}(F_1, \{B\})$

$C \to B$: ✓
$B \in \text{AttrHülle}(F_1, \{C\})$

$A \to C$: ✓
$C \in \text{AttrHülle}(F_1, \{A\})$

Äquivalenz von Mengen von FDs

Gegeben: Mengen F_1, F_2 von FDs.

Aufgabe: Zeige $F_1 \equiv F_2$ bzw. $F_1 \not\equiv F_2$?

$F_1 \equiv F_2$: Zeige $F_1 \subseteq F_2^+$ und $F_2 \subseteq F_1^+$

$F_1 \not\equiv F_2$: Finde FD $\alpha \to \beta \in F_1^+$ so dass $\alpha \to \beta \notin F_2^+$ (oder umgekehrt)

Zeige $\alpha \to \beta \notin F_2^+$: gdw. $\beta \notin \text{AttrHülle}(F_2, \alpha)$

Inhalt

Kanonische Überdeckung

Problem: gesucht ist eine möglichst knappe Darstellung von FDs ("Basis")

Lösung: die Kanonische Überdeckung

Mathematik: Basis eines Vektorraumes

Kanonische Überdeckung

Definition (Kanonische Überdeckung)

F_C heißt kanonische Überdeckung einer Menge von FDs F, wenn die folgenden drei Kriterien erfüllt sind:

1. $F_C^+ = F^+$ (F_C ist äquivalent zu F)
2. In F_C existieren keine FDs, die überflüssige Attribute enthalten
3. Jede linke Seite einer FD in F_C ist einzigartig.

Theorem

Zu jeder Menge F von FDs gibt es eine kanonische Überdeckung F_C.

Mathematik: Zu jeder Menge von Vektoren, die einen Vektorraum aufspannen, gibt es eine Basis.

Berechnung Kanonische Überdeckung (1)

Die Berechnung der Kanonischen Überdechung ergibt sich direkt aus der Definition:

1. Zerlege alle FDs mittels Dekomposition auf der rechten Seite (Äquivalenz gewährleistet)
2. Kürze überflüssige Attribute wie folgt:
   1. Führe für jede FD $\alpha \to B \in F$ die Linksreduktion durch, also: kann aus α ein Attribut gekürzt werden, sodass das Ergebnis äquivalent zur ursprünglichen Menge von FDs bleibt?
   2. Führe für jede (verbliebene) FD $\alpha \to B \in F$ die Rechtsreduktion durch, also: kann B gekürzt werden, sodass das Ergebnis äquivalent zur ursprünglichen Menge von FDs bleibt?
3. Fasse mittels der Vereinigungsregel FDs zusammen (Äquivalenz gewährleistet).

Damit in den Kürzungsschritten sichergestellt bleibt, dass nur äquivalente Ergebnisse erhalten werden, verwenden wir den Algorithmus AttrHülle

Berechnung Kanonische Überdeckung (2)

1. Zerlege alle FDs mittels Dekomposition auf der rechten Seite
2. Kürze überflüssige Attribute wie folgt:
   1. Führe für jede FD $\alpha \to B \in F$ die Linksreduktion durch:

      $\forall A \in \alpha$: gilt $B \in$ AttrHülle$(F, \alpha - A)$ (ist A überflüssig?)
      wenn ja: ersetze in F $\alpha \to B$ durch $(\alpha - A) \to B$.

   2. Führe für jede (verbliebene) FD $\alpha \to B \in F$ die Rechtsreduktion durch:

      gilt $B \in$ AttrHülle$(F - (\alpha \to B), \alpha)$ (ist B bzw. $\alpha \to B$ überflüssig?)
      wenn ja: streiche $\alpha \to B$
3. Fasse mittels der Vereinigungsregel FDs zusammen.

Begründung für ...

... die Linksreduktion:

$$B \in \text{AttrHülle}(F, \alpha - A) \Leftrightarrow F \equiv (F \setminus \{\alpha \to B\}) \cup \{(\alpha - A) \to B\}$$

$$(\text{wegen } B \in \text{AttrHülle}(F, \alpha - A) \Leftrightarrow F \models (\alpha - A) \to B)$$

... die Rechtsreduktion

$$B \in \text{AttrHülle}(F - (\alpha \to B), \alpha) \Leftrightarrow F \equiv F - (\alpha \to B)$$

$$(\text{wegen } B \in \text{AttrHülle}(F - (\alpha \to B), \alpha) \Leftrightarrow F - (\alpha \to B) \models \alpha \to B)$$

Berechnung Kanonische Überdeckung (Beispiel)

Beispiel

$F = \{A \rightarrow B, B \rightarrow C, AB \rightarrow C\}$

1. Dekomposition nicht notwendig
2. Kürzen
   1. Linksreduktion:

      $A \rightarrow B$: bereits reduziert

      $B \rightarrow C$: bereits reduziert

      $AB \rightarrow C$: $C \in \text{AttrHülle}(F, A)$ ja

      $\Rightarrow F := \{A \rightarrow B, B \rightarrow C, A \rightarrow C\}$
   2. Rechtsreduktion:

      $A \rightarrow B$: $B \in \text{AttrHülle}(F \setminus \{A \rightarrow B\}, A)$ nein

      $B \rightarrow C$: $C \in \text{AttrHülle}(F \setminus \{B \rightarrow C\}, B)$ nein

      $A \rightarrow C$: $C \in \text{AttrHülle}(F \setminus \{A \rightarrow C\}, A)$ ja

      $\Rightarrow F := \{A \rightarrow B, B \rightarrow C\}$
3. Vereinigungsregel nicht anwendbar $\Rightarrow F_C := \{A \rightarrow B, B \rightarrow C\}$

Kanonische Überdeckung

Beispiel

$F = \{A \to BD, AC \to E, CD \to E, E \to A, D \to C\}$

1. Dekomposition:
   $F = \{A \to B, A \to D, AC \to E, CD \to E, E \to A, D \to C\}$
2. Linksreduktion:

   $A \to B$: ok

   $A \to D$: ok

   $E \to A$: ok

   $D \to C$: ok

   $AC \to E$: $E \in \text{AttrHülle}(F, A)$ ja $\Rightarrow$
   $F = \{A \to B, A \to D, A \to E, CD \to E, E \to A, D \to C\}$

   $CD \to E$: $E \in \text{AttrHülle}(F, C)$ nein
   $E \in \text{AttrHülle}(F, D)$ ja $\Rightarrow$
   $F = \{A \to B, A \to D, A \to E, D \to E, E \to A, D \to C\}$

Kanonische Überdeckung

Beispiel

$F = \{A \to B, A \to D, A \to E, D \to E, E \to A, D \to C\}$

3. Rechtsreduktion:

$$\begin{aligned}
A \to B&: B \in \text{AttrHülle}(F \setminus \{A \to B\}, A) \text{ nein}\\
A \to D&: D \in \text{AttrHülle}(F \setminus \{A \to D\}, A) \text{ nein}\\
A \to E&: E \in \text{AttrHülle}(F \setminus \{A \to E\}, A) \text{ ja} \Rightarrow\\
& F = \{A \to B, A \to D, D \to E, E \to A, D \to C\}\\
D \to E&: E \in \text{AttrHülle}(F \setminus \{D \to E\}, D) \text{ nein}\\
E \to A&: A \in \text{AttrHülle}(F \setminus \{E \to A\}, E) \text{ nein}\\
D \to C&: C \in \text{AttrHülle}(F \setminus \{D \to C\}, D) \text{ nein}
\end{aligned}$$

4. Zusammenfassen: $F_C = \{A \to BD, D \to EC, E \to A\}$

Inhalt

Schlüssel

Definition (Schlüssel)

$\gamma \subseteq \mathcal{R}$ ist ein Kandidatenschlüssel oder Schlüssel, falls die beiden folgenden Bedingungen erfüllt sind:

1. $\gamma \rightarrow \mathcal{R}$
2. γ ist minimal, d.h. für alle $A \in \gamma$: $(\gamma - \{A\}) \not\rightarrow \mathcal{R}$

Im Relationenmodell: Auszeichnung eines Schlüssels als Primärschlüssel zur Verknüpfung von Tabellen mittels Primär- und Fremdschlüssel

Definition (Superschlüssel)

$\gamma \subseteq \mathcal{R}$ ist ein Superschlüssel (Oberschlüssel), falls $\gamma \rightarrow \mathcal{R}$

- keine Minimalität bei Superschlüsseln
- ein Superschlüssel ist eine Obermenge eines Schlüssels

$\Rightarrow$ jeder Schlüssel ist auch ein Superschlüssel

Schlüsselbestimmung

Beispiel

Eine Stadt wird beschrieben durch ihren Namen (Name), das Bundesland (BLand) in dem sie liegt, die Vorwahl (VW) und die Einwohnerzahl (EW).
Frage: Welche FDs gelten in diesem Szenario?

- Ortsnamen sind innerhalb eines Bundeslandes eindeutig d.h. $\{\text{Name, BLand}\} \rightarrow \{\text{VW, EW}\}$
- Mehrere Städte können dieselbe Vorwahl haben, aber nur, wenn sie einen unterschiedlichen Namen haben d.h. $\{\text{Name, VW}\} \rightarrow \{\text{BLand, EW}\}$

Frage: Welche Kandidatenschlüssel ergeben sich aufgrund der gefundenen FDs?

- $\{\text{Name, BLand}\}$
- $\{\text{Name, VW}\}$

Schlüsselbestimmung mittels AttrHülle

Problem: Suche nach Kandidatenschlüsseln einer Relation R aufgrund der vorhandenen FDs.

Definition (Schlüssel - Wiederholung)

$\gamma \subseteq \mathcal{R}$ ist ein Kandidatenschlüssel oder Schlüssel, falls die beiden folgenden Bedingungen erfüllt sind:

1. $\gamma \rightarrow \mathcal{R}$
2. γ ist minimal, d.h. $\forall A \in \gamma : (\gamma - \{A\}) \not\rightarrow \mathcal{R}$

1. $\gamma \rightarrow \mathcal{R}$, wenn $\mathrm{AttrHülle}(F, \gamma) = \mathcal{R}$
2. die Minimalität ist erfüllt, wenn für jedes Attribut A aus γ gilt: $\mathrm{AttrHülle}(F, \gamma - \{A\}) \neq \mathcal{R}$

Schlüsselbestimmung mittels AttrHülle

Beispiel

$\mathcal{R} = \{ABCDEF\}, F_d = \{C \rightarrow BDAE\}$

- Zeitraubende Vorgehensweise: Durchprobieren aller einelementigen, aller zweielementigen, aller dreielementigen Schlüsselkandidaten mittels AttrHülle
- Alternativ: Verwendung der folgenden Heuristik: Alle Attribute, die nicht auf der rechten Seite vorkommen, können mittels AttrHülle nicht hergeleitet werden und müssen daher im Schlüssel enthalten sein.
- Hier: C und F kommen rechts nicht vor, daher folgender Versuch: AttrHülle$(\{C \rightarrow BDAE\}, CF) = \{C, F, B, D, A, E\}$

$$\Rightarrow$$

CF ist Schlüssel von $\mathcal{R}$

Schlüsselbestimmung mittels AttrHülle

Beispiel

$\mathcal{R} = \{ABCDEF\}, F_d = \{C \to BD, D \to AE, E \to CF, F \to E\}$
Heuristik hier nicht zielführend, da alle Attribute hergeleitet werden können.

- C: AttrHülle$(F_d, C) = \{C, B, D, A, E, F\} \Rightarrow C$ ist Schlüssel von $\mathcal{R}$
  Achtung: C kann aus E hergeleitet werden also:
- E: AttrHülle$(F_d, E) = \{E, C, B, D, A, F\} \Rightarrow E$ ist auch Schlüssel von $\mathcal{R}$
  Achtung: E kann aus D oder F hergeleitet werden also:
- D: AttrHülle$(F_d, D) = \{D, A, E, C, F, B\} \Rightarrow D$ ist Schlüssel von $\mathcal{R}$,
- F: AttrHülle$(F_d, F) = \{F, E, C, B, D, A\} \Rightarrow F$ ist Schlüssel von $\mathcal{R}$

$$\Rightarrow$$

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

Algorithmus zur Berechnung aller Schlüssel

Input: $(F, \mathcal{R})$: Menge F von FDs und Schema $\mathcal{R}$

Output: Menge aller Schlüssel von $\mathcal{R}$

AlleSchlüssel $(F, \mathcal{R})$

$keys = \{minimize(F, \mathcal{R}, \mathcal{R})\}$ // finde 1. Schlüssel
for each $key \in keys$:
  for each $att \in key$:
    for each $\alpha \to \beta \in F$:
      if $att \in \beta$:
        $nkey = (key \setminus \{att\}) \cup \alpha$
        if $\nexists k \in keys$ mit $k \subseteq nkey$:
          $keys = keys \cup \{minimize(F, nkey, \mathcal{R})\}$
return keys

$minimize(F, \gamma, \delta)$: gibt eine minimale Teilmenge $\gamma' \subseteq \gamma$ zurück so dass $\delta \subseteq$ AttrHülle(F, γ').

Berechnung aller Schlüssel eines Schemas

Beispiel

$\mathcal{R} = \{ABCDEFG\}$, $F_d = \{B \rightarrow BEF, DE \rightarrow AC, A \rightarrow BDG\}$

1. Finde 1. Schlüssel: AttrHülle$(F_d, \{A\}) = \mathcal{R} \Rightarrow keys = \{A\}$
2. Konstruiere neue Schlüssel aus $keys$:
   - $key = A$, $att = A$:
     - $B \rightarrow BEF$, $A \rightarrow BDG$: $A \notin \beta$
     - $DE \rightarrow AC$: $nkey = DE$
       $minimize$: AttrHülle$(F_d, \{D\}) \neq \mathcal{R}$ und AttrHülle$(F_d, \{E\}) \neq \mathcal{R}$
       $\Rightarrow keys = \{A, DE\}$
   - $key = DE$, $att = D$:
     - $B \rightarrow BEF$, $DE \rightarrow AC$: $D \notin \beta$
     - $A \rightarrow BDG$: $nkey = AE$; $A \subseteq AE \Rightarrow keys = \{A, DE\}$
   - $key = DE$, $att = E$:
     - $A \rightarrow BDG$, $DE \rightarrow AC$: $E \notin \beta$
     - $B \rightarrow BEF$: $nkey = BD$;
       $minimize$: AttrHülle$(F_d, \{B\}) \neq \mathcal{R}$ und AttrHülle$(F_d, \{D\}) \neq \mathcal{R}$
       $\Rightarrow keys = \{A, DE, BD\}$

Berechnung aller Schlüssel eines Schemas

Beispiel (Fortsetzung)

$\mathcal{R} = \{ABCDEFG\}$, $F_d = \{B \to BEF, DE \to AC, A \to BDG\}$

2. Konstruiere neue Schlüssel aus *keys*:
   - $key = BD$, $att = B$:
     - $DE \to AC$: $B \notin \beta$
     - $A \to BDG$: $nkey = AD$; $A \subseteq AD \Rightarrow keys = \{A, DE, BD\}$
     - $B \to BEF$: $nkey = BD$; $BD \subseteq BD \Rightarrow keys = \{A, DE, BD\}$
   - $key = BD$, $att = D$:
     - $DE \to AC$, $B \to BEF$: $D \notin \beta$
     - $A \to BDG$: $nkey = AB$; $A \subseteq AB \Rightarrow keys = \{A, DE, BD\}$
3. $keys = \{A, DE, BD\}$

Lernziele

- Was sind FDs?
  - Wann ist eine FD erfüllt, wie kann man das überprüfen?
- Was ist die Attributhülle/Hülle von FDs?
  - Wie kann man sie berechnen?
- Wann sind zwei Mengen von FDs äquivalent?
- Was sind die Armstrong Axiome?
  - Wozu braucht man sie, welche gibt es?
- Was ist die kanonische Überdeckung?
  - Wie kann man sie berechnen?
- Was sind (Super)Schlüssel?
  - Wie kann ich sie erkennen/überprüfen?