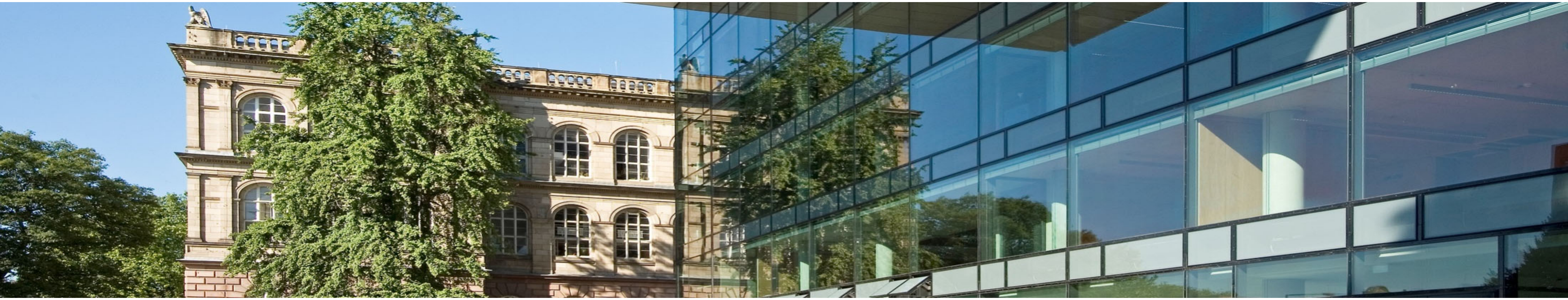


6. Relationale Entwurfstheorie

1. Funktionale Abhängigkeiten
2. Armstrong-Kalkül
3. Zerlegung von Relationen
4. Normalformen und Normalisierungen

Informatik 5 -
Informationssysteme
und Datenbanken
RWTH AACHEN UNIVERSITY

Einführung

Bis jetzt:

- Nutzen- und Anforderungsanalys (Pflichtenheft)
- Entity-Relationship-Entwurf
- relationales Schema

Zu tun:

- Feintuning des erstellten Schemas auf der Basis von intrarelationalen Abhängigkeiten
  - Funktionale Abhängigkeiten
  - Kriterien für gute Schemata, schlechte Schemata
  - Normalformen
  - Algorithmen zur Normalisierung

Was ist faul mit diesem Schema?

ProfVorl						
PersNr	Name	Rang	Raum	VorlNr	Titel	SWS
2125	Sokrates	W3	226	5041	Ethik	4
2125	Sokrates	W3	226	5049	Mäeutik	2
2125	Sokrates	W3	226	4052	Logik	4
...	...	...	...	...	...	...
2132	Popper	W2	52	5259	Der Wiener Kreis	2
2137	Kant	W3	7	4630	Die 3 Kritiken	4

Änderungs-Anomalie

ProfVorl						
PersNr	Name	Rang	Raum	VorlNr	Titel	SWS
2125	Sokrates	W3	226	5041	Ethik	4
2125	Sokrates	W3	226	5049	Mäeutik	2
2125	Sokrates	W3	226	4052	Logik	4
...	...	...	...	...	...	...
2132	Popper	W2	52	5259	Der Wiener Kreis	2
2137	Kant	W3	7	4630	Die 3 Kritiken	4

- Angenommen Sokrates soll von Raum 226 nach Raum 233 umziehen
- Die Information ‘Raum’ existiert in diesem Fall mehrfach
- Lösung: Änderung aller Einträge gleichzeitig
  – hoher Speicherbedarf durch Redundanz
  – erhöhter Zeitbedarf bei Änderungen

Einfüge-Anomalie

ProfVorl						
PersNr	Name	Rang	Raum	VorlNr	Titel	SWS
2125	Sokrates	W3	226	5041	Ethik	4
2125	Sokrates	W3	226	5049	Mäeutik	2
2125	Sokrates	W3	226	4052	Logik	4
...	...	...	...	...	...	...
2132	Popper	W2	52	5259	Der Wiener Kreis	2
2137	Kant	W3	7	4630	Die 3 Kritiken	4

- Schema kombiniert Informationen verschiedener unpassender Entitytypen
  - Hinzufügen eines Professors ohne Vorlesung

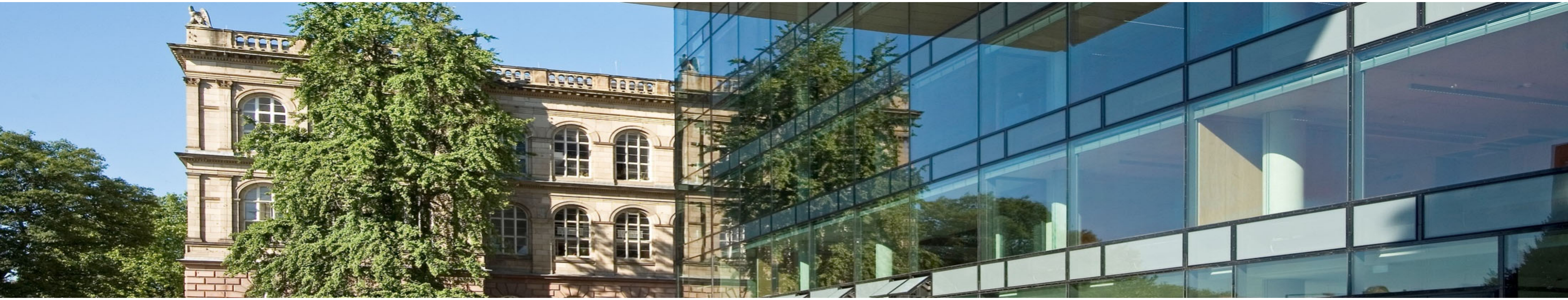
    ⇒ NULL-Werte in VorlNr, Titel und SWS
  - Analog: Hinzufügen einer Vorlesung zu der noch kein Dozent festgelegt wurde

    ⇒ NULL-Werte in PersNr, Name, Rang und Raum

Lösch-Anomalie

ProfVorl						
PersNr	Name	Rang	Raum	VorlNr	Titel	SWS
2125	Sokrates	W3	226	5041	Ethik	4
2125	Sokrates	W3	226	5049	Mäeutik	2
2125	Sokrates	W3	226	4052	Logik	4
…	…	…	…	…	…	…
2132	Popper	W2	52	5259	Der Wiener Kreis	2
2137	Kant	W3	7	4630	Die 3 Kritiken	4

- Schema kombiniert Informationen verschiedener unpassender Entitytypen
  - Löschen von Elementen eines Entitytyps kann Verlust eines anderen Entitytyps bewirken
    - Löschen des Eintrags zu „Der Wiener Kreis“ (die einzige Vorlesung von Popper) würde auch Informationen zu Popper löschen
    - Alternative: Prüfen der *gesamten* Datenbank, ob dieser Eintrag die einzige Vorlesung von Popper ist. In diesem Fall durch NULL-Werte ersetzen



6. Relationale Entwurfstheorie

1. **Funktionale Abhängigkeiten**
2. Armstrong-Kalkül
3. Zerlegung von Relationen
4. Normalformen und Normalisierungen

Informatik 5 - Informationssysteme und Datenbanken

RWTH AACHEN UNIVERSITY

Funktionale Abhängigkeiten

- Das zentrale Konzept der relationalen Entwurfstheorie

- Sei X die Attributmenge eines Relationenschemas $\mathcal{R}$. Die funktionalen Abhängigkeiten über X bilden eine zweistellige Relation „→“ auf den Attributmengen aus X:

$$\alpha \rightarrow \beta, \qquad \text{für } \alpha, \beta \subseteq X.$$

(gesprochen: von Alpha nach Beta)

- In Worten: β ist funktional abhängig von α
  oder die α-Werte bestimmen die β-Werte funktional (d.h. eindeutig)

- Für zwei Attributmengen $\alpha, \beta \subseteq X$ und eine Relation R sagen wir R *erfüllt* die funktionale Abhängigkeit $\alpha \rightarrow \beta$, wenn gilt:

$$r.\alpha = t.\alpha \ \text{ impliziert } \ r.\beta = t.\beta \ \text{ für alle } r, t \in R.$$

Funktionale Abhängigkeiten

- Verallgemeinerung der Schlüsseleigenschaft
  - Eindeutigkeitseigenschaft der Schlüssel als funktionale Abhängigkeit:

$$\alpha \rightarrow X$$

- Eine funktionale Abhängigkeit $\alpha \rightarrow \beta$ lässt sich ebenfalls als intrarelationale Abhängigkeit auffassen:

$$\sigma_{\alpha\rightarrow\beta}\colon \mathrm{Rel}(X) \rightarrow \{true, false\}, \qquad R \longmapsto \begin{cases} true, & \text{falls } \alpha \rightarrow \beta \text{ in } R \text{ gilt} \\ false, & \text{sonst} \end{cases}$$

- Ist $\beta \subseteq \alpha$, so heißt $\alpha \rightarrow \beta$ eine *triviale Abhängigkeit*

- Funktionale Abhängigkeiten werden auch als FDs (functional dependencies) abgekürzt

Funktionale Abhängigkeiten – Beispiel

- Betrachte Schema $\mathcal{R}$ mit Attributmenge $\{A, B, C, D\}$ und FD $\{A\} \rightarrow \{B\}$.
  Die Ausprägung r erfüllt diese FD:
  nur für die Tupel t_2, t_3 gilt $t_2.A = t_3.A\ (= a_1)$
  und für diese gilt ebenfalls $t_2.B = t_3.B\ (= b_1)$
  - diese Ausprägung erfüllt auch die FDs
    - $\{A\} \rightarrow \{C\}$
    - $\{A, B\} \rightarrow \{C\}$
    - $\{C, D\} \rightarrow \{B\}$

    nicht aber die FDs
    - $\{B\} \rightarrow \{C\}$
    - $\{A, B, C\} \rightarrow \{D\}$

	r			
	A	B	C	D
t_1	a_4	b_2	c_4	d_3
t_2	a_1	b_1	c_1	d_1
t_3	a_1	b_1	c_1	d_2
t_4	a_2	b_2	c_3	d_2
t_5	a_3	b_2	c_4	d_3

Funktionale Abhängigkeiten – Beispiel

- Wichtig:
  - funktionale Abhängigkeiten beschreiben die Menge aller gültigen Relationen (wenn nichts anderes gesagt wird)
  - üblicherweise wird gefragt: welche zusätzlichen FDs lassen sich aus den gegebenen FDs ableiten
  - es wird nicht gefragt: welche zusätzlichen FDs erfüllt diese konkrete Ausprägung

	r			
	A	B	C	D
t_1	a_4	b_2	c_4	d_3
t_2	a_1	b_1	c_1	d_1
t_3	a_1	b_1	c_1	d_2
t_4	a_2	b_2	c_3	d_2
t_5	a_3	b_2	c_4	d_3

- Übrigens: die Notation

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

wird häufig auch abgekürzt, z.B. durch

$$A, B, C \rightarrow D.$$

Funktionale Abhängigkeiten – Beispiel

Student			
$\underline{\text{MatrNr}: Inte}$	$\text{Name}: String$	$\text{\#Semester}: Int$	$\text{Status}: Status$
1234	Michael	6	eingeschrieben
5678	Andrea	4	eingeschrieben
4711	Sabine	8	beurlaubt
815	Franz	12	exmatrikuliert

$\sigma_{\alpha \to \beta}$	Wert in der Ausprägung	Wert in allen Ausprägungen
$MatrNr \to Name, Semester, Status$		
$Name \to Semester, Status, MatrNr$		
$Semester \to Status$		
$Status \to Semester$		
$MatrNr, Name \to Semester, Status$		

Funktionale Abhängigkeiten – Beispiel

Student			
$\underline{\mathbf{MatrNr}: \boldsymbol{Inte}}$	$\mathbf{Name}: \boldsymbol{String}$	$\mathbf{\#Semester}: \boldsymbol{Int}$	$\mathbf{Status}: \boldsymbol{Status}$
1234	Michael	6	eingeschrieben
5678	Andrea	4	eingeschrieben
4711	Sabine	8	beurlaubt
815	Franz	12	exmatrikuliert

$\sigma_{\alpha \to \beta}$	Wert in der Ausprägung	Wert in allen Ausprägungen
$MatrNr \to Name, Semester, Status$	true	true
$Name \to Semester, Status, MatrNr$	true	false
$Semester \to Status$	true	false
$Status \to Semester$	false	false
$MatrNr, Name \to Semester, Status$	true	true

Überprüfen funktionaler Abhängigkeiten

- Ein einfacher Algorithmus zur Überprüfung einer FD:
  - Eingabe: eine Relation R und eine FD $\alpha \rightarrow \beta$
  - Ausgabe: *ja* genau dann, wenn $\alpha \rightarrow \beta$ in R erfüllt ist
  - Algorithmus:
    - sortiere R nach den α-Werten
    - falls alle Gruppen, bestehend aus Tupeln mit gleichen α-Werten, auch gleiche β-Werte aufweisen: *ja*; sonst: *nein*

- Die Laufzeit dieses Algorithmus wird durch die Sortierung dominiert
  - Komplexität $O\big(n \log(n)\big)$

Schlüssel

- Präzisierung des Schlüsselbegriffs
  - Dazu sei $\mathcal{R}$ ein Relationenschema mit Attributmenge X und funktionalen Abhängigkeiten F
- Superschlüssel (Eindeutigkeit) :
  - $\alpha \subseteq X$ heißt *Superschlüssel*, falls $\alpha \rightarrow X$ gilt
  - α bestimmt also alle anderen Attributwerte
  - X selbst ist stets auch ein Superschlüssel, da trivialerweise $X \rightarrow X$ gilt
- voll funktional abhängig (Minimalität):
  - $\beta \subseteq X$ heißt *voll funktional abhängig* von α, falls
    - $\alpha \rightarrow \beta$ gilt
    - $\alpha - \{A\} \nrightarrow \beta$ für alle $A \in \alpha$ gilt, d.h. α kann nicht „verkleinert“ werden

Schlüssel

- Präzisierung des Schlüsselbegriffs
  - Dazu sei $\mathcal{R}$ ein Relationenschema mit Attributmenge X und funktionalen Abhängigkeiten F
- Schlüsselkandidat:
  - Eine Attributmenge $\alpha \subseteq X$ heißt *Schlüsselkandidat*, falls X voll funktional abhängig von α ist
- Primärschlüssel:
  - In einem Relationenschema wird einer der Schlüsselkandidaten als *Primärschlüssel* ausgewählt
  - Fremdschlüssel sollten z.B. immer nur auf den Primärschlüssel verweisen

Schlüssel – Beispiel

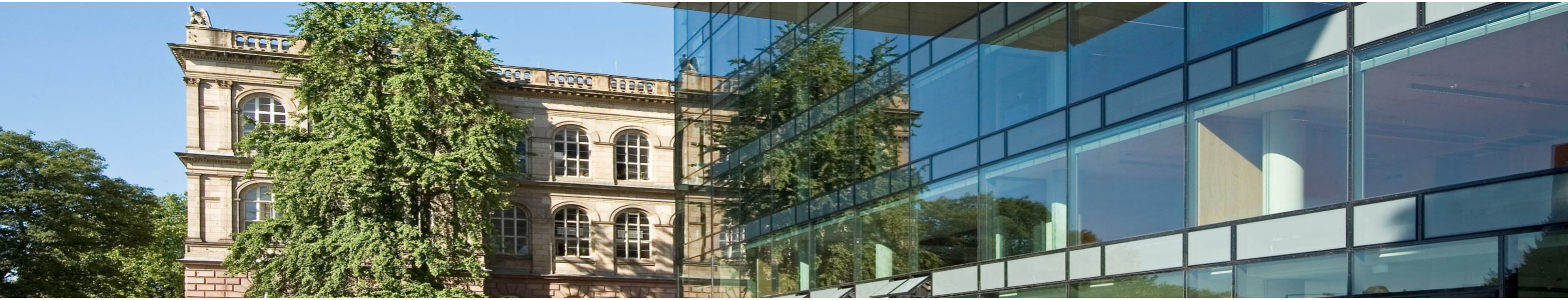
Orte			
Name	BLand	Vorwahl	EW
Frankfurt	Hessen	069	690.000
Frankfurt	Brandenburg	0335	60.000
München	Bayern	089	1.378.000.000
Passau	Bayern	0851	50.000

- Annahme: Ortsnamen sind innerhalb eines Bundeslandes eindeutig
- Schlüsselkandidaten:

$$\{\text{Name}, \text{BLand}\}, \qquad \{\text{Name}, \text{Vorwahl}\}$$

beide sind minimal:

  - Städte in unterschiedlichen Bundesländern können denselben Namen besitzen
  - kleine Dörfer können sich dieselbe Vorwahl teilen



6. Relationale Entwurfstheorie

1. Funktionale Abhängigkeiten
2. **Armstrong-Kalkül**
3. Zerlegung von Relationen
4. Normalformen und Normalisierungen

Informatik 5 -
Informationssysteme
und Datenbanken
RWTH AACHEN UNIVERSITY

Bestimmung aller funktionalen Abhängigkeiten

- Frage: Ausgehend von einer Menge funktionaler Abhängigkeiten F (beim Datenbank-Entwurf erstellt), welche zusätzlichen funktionalen Abhängigkeiten sind implizit immer erfüllt?

- *Beispiel*:
  - Erweiterung von Universitäts-Beispiel um Adressen
  - erster Entwurf für Professoren und Adressen:

$$\text{ProfessorenAdressen(PersNr, Name, Rang, Raum, Ort, Straße, Hausnummer, PLZ, Vorwahl, Bundesland)}$$

Bestimmung aller funktionalen Abhängigkeiten – Beispiel

$\text{ProfessorenAdressen(PersNr, Name, Rang, Raum, Ort, Straße, Hausnummer, PLZ, Vorwahl, Bundesland)}$

- Funktionale Abhängigkeiten F:
  - $\text{PersNr} \rightarrow \text{PersNr, Name, Rang, … , Vorwahl, Bundesland}$
  - $\text{Raum} \rightarrow \text{PersNr}$
  - $\text{Ort, Bundesland} \rightarrow \text{Vorwahl}$
  - $\text{Ort, Bundesland, Straße, Hausnummer} \rightarrow \text{PLZ}$
  - $\text{PLZ} \rightarrow \text{Ort, Bundesland}$
- implizierte funktionale Abhängigkeiten:
  - $\text{Raum} \rightarrow \text{PersNr, Name, Rang, … , Vorwahl, Bundesland}$
  - $\text{PLZ} \rightarrow \text{Vorwahl}$
  - …

Implizierte funktionale Abhängigkeiten (Formalisierung)

- Seien X eine Attributmenge und F eine Menge von funktionalen Abhängigkeiten über X.
- Semantische Grundlagen:
  - Eine Relationsausprägung R *erfüllt* F, falls R jede funktionale Abhängigkeit $f \in F$ erfüllt (man sagt auch: R ist ein Modell von F)
  - Schreibweise:

$$R \vDash F \text{ oder } R \vDash f$$

  - Die Menge aller *gültigen Ausprägungen* des Schemas $\mathcal{R}$ ist

$$\mathrm{Sat}(F) = \{\, R \in \mathrm{Rel}(X) \mid R \vDash F \,\}$$

  - Zwei Mengen F, G von funktionalen Abhängigkeiten über X heißen *äquivalent*, falls sie die gleichen gültigen Relationen definieren:

$$\mathrm{Sat}(F) = \mathrm{Sat}(G)$$

Implizierte funktionale Abhängigkeiten (Definition)

- Seien X eine Attributmenge und F eine Menge von funktionalen Abhängigkeiten über X.
- *Implikation*:
  - Wir sagen, dass die Menge funktionaler Abhängigkeiten F die funktionale Abhängigkeit f impliziert, falls alle F erfüllenden Relationenausprägungen $R \in \mathrm{Sat}(F)$ auch f erfüllen.
  - Schreibweise:

$$F \vDash f$$

$F \vDash f$ ist eine semantische Beziehung

- nicht praktikabel: Überprüfe jede gültige Ausprägung $R \in \mathrm{Sat}(F)$, ob f gilt
- stattdessen: Armstrong-Kalkül

Armstrong-Kalkül

- Algorithmische Bestimmung aller implizierten funktionalen Abhängigkeiten

- Hilfsmittel: Armstrong-Axiome

  Seien $\alpha, \beta, \gamma \subseteq X$ Attributmengen.

  – Reflexivität (A_1): Ist $\beta \subseteq \alpha$ eine Teilmenge von α, so gilt auch $\alpha \rightarrow \beta$.

  – Verstärkung (A_2): Falls $\alpha \rightarrow \beta$ gilt, so gilt auch $\alpha\gamma \rightarrow \beta\gamma$. Wobei hier $\alpha\gamma \coloneqq \alpha \cup \gamma$.

  – Transitivität (A_3): Falls $\alpha \rightarrow \beta$ und $\beta \rightarrow \gamma$ gilt, so gilt auch $\alpha \rightarrow \gamma$.

Armstrong-Kalkül

- Algorithmische Bestimmung aller implizierten funktionalen Abhängigkeiten

- *Ableitbar* :
  - Wir sagen, dass die funktionale Abhängigkeit f aus der Menge der funktionalen Abhängigkeiten F *ableitbar* ist, falls:
    Es gibt eine endliche Folge $f_1, \ldots, f_{n-1}, f_n = f$, sodass für jedes $1 \leq i \leq n$ gilt:

    f_i erhält man aus $F \cup \{f_1, \ldots, f_{i-1}\}$ durch Anwendung der Axiome A_1, A_2 oder A_3

  - Schreibweise

$$F \vdash f$$

$F \vdash f$ ist eine syntaktische Beziehung

Armstrong-Kalkül – Beispiel

- *Ableitbar* :

  $\text{ProfessorenAdressen(PersNr, Name, Rang, Raum, } \qquad \text{Ort, Straße, Hausnummer, PLZ, Vorwahl, Bundesland)}$

- Funktionale Abhängigkeiten F:
  - $\text{Ort, Bundesland} \rightarrow \text{Vorwahl}$
  - $\text{PLZ} \rightarrow \text{Ort, Bundesland}$
  - ...
- Mithilfe der Transitivitätsregel (A_3) :Falls $\alpha \rightarrow \beta$ und $\beta \rightarrow \gamma$ gilt, so gilt auch $\alpha \rightarrow \gamma$ erhält man aus

$$\text{PLZ} \rightarrow \text{Ort, Bundesland,}$$

$$\text{Ort, Bundesland} \rightarrow \text{Vorwahl}$$

die funktionale Abhängigkeit $\text{PLZ} \rightarrow \text{Vorwahl.}$

$$F \vdash (\text{PLZ} \rightarrow \text{Vorwahl})$$

Armstrong-Kalkül – Korrektheit und Vollständigkeit

- Liefert das Armstrong-Kalkül alle „gültigen“ bzw. von F implizierten FDs?

  Sei X eine Attributmenge und F eine Menge von FDs über X

  – Zu F nennen wir $F^+ = \{ f \mid F \vdash f \}$ die *(geschlossene) Hülle* von F

  – Gilt also

$$F^+ = \{ f \mid F \vDash f \} ?$$

- Ja, der Armstrong-Kalkül ist *korrekt* und *vollständig*.

  – *korrekt*: Für jede funktionale Abhängigkeit f mit $F \vdash f$ gilt auch $F \vDash f$

  (es lassen sich nur „gültige“ funktionale Abhängigkeiten ableiten, „⊆“)

  – *vollständig*: Jede von F implizierte funktionale Abhängigkeit f (also $F \vDash f$) lässt sich mithilfe des Armstrong-Kalküls ableiten, d.h. $F \vdash f$ („⊇“)

  (Beweis der Korrektheit jetzt, Beweis der Vollständigkeit später)

Armstrong-Kalkül – Korrektheit

- Korrektheit der Reflexivitätsregel (A_1): Für $\beta \subseteq \alpha$ gilt $\alpha \rightarrow \beta$.
  - Seien $R \in \mathrm{Sat}(F)$ eine gültige Relationsausprägung, $\alpha \subseteq X$ und $\beta \subseteq \alpha$. Außerdem seien $t_1, t_2 \in R$ zwei beliebige Tupel mit $t_1[\alpha] = t_2[\alpha]$. Dann gilt auch $t_1[\beta] = t_2[\beta]$.

    Insgesamt: $\alpha \rightarrow \beta$ gilt auch in R.

Armstrong-Kalkül – Korrektheit

- Korrektheit der Verstärkungsregel (A_2): Für $\alpha \rightarrow \beta$ gilt auch $\alpha\gamma \rightarrow \beta\gamma$.
  - Seien $R \in \mathrm{Sat}(F)$ eine gültige Relationsausprägung, $\alpha, \beta, \gamma \subseteq X$ und $\alpha \rightarrow \beta \in F$. Außerdem seien $t_1, t_2 \in R$ zwei beliebige Tupel mit $t_1[\alpha \cup \gamma] = t_2[\alpha \cup \gamma]$. Dann gilt auch $t_1[\beta] = t_2[\beta]$ und $t_1[\gamma] = t_2[\gamma]$. Daraus folgt $t_1[\beta \cup \gamma] = \mathrm{t}_2[\beta \cup \gamma]$. Insgesamt: $\alpha \cup \gamma \rightarrow \beta \cup \gamma$ gilt auch in R.

Armstrong-Kalkül – Korrektheit

- Korrektheit der Transitivitätsregel (A_3): Für $\alpha \to \beta, \beta \to \gamma$ gilt auch $\alpha \to \gamma$
  - Seien $R \in \mathrm{Sat}(F)$ eine gültige Relationenausprägung, $\alpha, \beta, \gamma \subseteq X$ und $\alpha \to \beta,\ \beta \to \gamma \in F$. Außerdem seien $t_1, t_2 \in R$ zwei beliebige Tupel mit $t_1[\alpha] = t_2[\alpha]$. Wegen $\alpha \to \beta$, gilt also auch $t_1[\beta] = t_2[\beta]$. Wegen $\beta \to \gamma$, gilt dann auch $t_1[\gamma] = t_2[\gamma]$. Insgesamt: $\alpha \to \gamma$ gilt auch in R.

Armstrong-Kalkül – Erweiterung

Es ist für den Herleitungsprozess komfortabel, weitere Regeln hinzuzunehmen

- Erweiterung der Armstrong-Axiome um drei Regeln:

  Seien $\alpha, \beta, \gamma, \delta \subseteq X$ Attributmengen.

  - Vereinigung (A_4): Gelten $\alpha \rightarrow \beta$ und $\alpha \rightarrow \gamma$, so gilt auch $\alpha \rightarrow \beta\gamma$.
  - Dekomposition (A_5): Falls $\alpha \rightarrow \beta\gamma$ gilt, so gelten auch $\alpha \rightarrow \beta$ und $\alpha \rightarrow \gamma$.
  - Pseudotransitivität (A_6): Falls $\alpha \rightarrow \beta$ und $\beta\gamma \rightarrow \delta$ gilt, so gilt auch $\alpha\gamma \rightarrow \delta$.

Armstrong-Kalkül – Erweiterung

- Ableitung der Vereinigungsregel (A_4) :Gelten $\alpha{\rightarrow}\beta$ und $\alpha{\rightarrow}\gamma$, so gilt auch $\alpha{\rightarrow}\beta\gamma$.
  - Seien $\alpha \rightarrow \beta$ und $\alpha \rightarrow \gamma \in F$. Über die Grundregeln erhalten wir

$$(A_2) : \text{Da } \alpha \rightarrow \beta \text{ gilt } \alpha\gamma \rightarrow \beta\gamma$$

$$(A_2) : \text{Da } \alpha \rightarrow \gamma \text{ gilt } \alpha \rightarrow \alpha\gamma$$

$$(A_3) : \text{Da } \alpha \rightarrow \alpha\gamma \text{ und } \alpha\gamma \rightarrow \beta\gamma \text{ gilt } \alpha \rightarrow \beta\gamma$$

Armstrong-Kalkül – Erweiterung

- Ableitung der Dekompositionsregel ($A_{5)}$):

  Falls $\alpha \rightarrow \beta\gamma$ gilt, so gelten auch $\alpha \rightarrow \beta$ und $\alpha \rightarrow \gamma$.

  – Sei $\alpha \rightarrow \beta\gamma \in F$. Über die Grundregeln erhalten wir

$$(A_1) : \text{Da } \beta\gamma \coloneqq \beta \cup \gamma \text{ gilt } \beta\gamma \rightarrow \beta$$

$$(A_1) : \text{Da } \beta\gamma \coloneqq \beta \cup \gamma \text{ gilt } \beta\gamma \rightarrow \gamma$$

$$(A_3) : \text{Da } \alpha \rightarrow \beta\gamma \text{ und } \beta\gamma \rightarrow \beta \text{ gilt } \alpha \rightarrow \beta$$

$$(A_3) : \text{Da } \alpha \rightarrow \beta\gamma \text{ und } \beta\gamma \rightarrow \gamma \text{ gilt } \alpha \rightarrow \gamma$$

Armstrong-Kalkül – Erweiterung

- Ableitung der Pseudotransitivitätsregel (A_6) :

  Falls $\alpha \rightarrow \beta$ und $\beta\gamma \rightarrow \delta$ gilt, so gilt auch $\alpha\gamma \rightarrow \delta$

  Sei $\alpha \rightarrow \beta$ und $\beta\gamma \rightarrow \delta \in F$. Über die Grundregeln erhalten wir

$$(A_2): \text{Da } \alpha \rightarrow \beta \text{ gilt } \alpha\gamma \rightarrow \beta\gamma$$

$$(A_3): \text{Da } \alpha\gamma \rightarrow \beta\gamma \text{ und } \beta\gamma \rightarrow \delta \text{ gilt } \alpha\gamma \rightarrow \delta$$

Armstrong-Kalkül – Vereinigung

$$\text{ProfessorenAdressen(PersNr, Name, Rang, Raum, Ort, Straße, Hausnummer, PLZ, Vorwahl, Bundesland)}$$

- Funktionale Abhängigkeiten F:
  - $\text{Ort, Bundesland} \rightarrow \text{Vorwahl}$
  - $\text{Ort, Bundesland, Straße, Hausnummer} \rightarrow \text{PLZ}$
  - …

- Beispiel: $f = \text{Ort, Bundesland, Straße, Hausnummer} \rightarrow \text{PLZ, Vorwahl}$ ist ebenfalls aus F ableitbar
  - $(A_1) : f_1 = (\text{Ort, Bundesland, Straße, Hausnummer} \rightarrow \text{Ort, Bundesland})$
  - $(A_3) : f_2 = (\text{Ort, Bundesland, Straße, Hausnummer} \rightarrow \text{Vorwahl})$
  - $(\boldsymbol{A_4}) : \boldsymbol{f} = \boldsymbol{f_3} = (\textbf{Ort, Bundesland, Strasse, Hausnummer} \rightarrow \textbf{PLZ, Vorwahl})$

Attributhülle

- Oft ist man nicht an der gesamten Hülle F^+ interessiert:
  - Welche Attribute sind unter einer gegebenen Menge von FDs F von einer bestimmten Attributmenge α funktional bestimmt?
  - Man nennt α^+ die Attributhülle von α unter F

$$\alpha^+ = \{\, x \mid \text{es gibt } \alpha \to \beta \in F^+ \text{ mit } x \in \beta \}$$

  - Algorithmus zur Bestimmung von $\mathrm{AttrHülle}(F, \alpha)$:
    - $Erg \ := \alpha$

      $\mathbf{while}$ (Änderungen an Erg) $\mathbf{do}$

      $\quad\mathbf{foreach}$ FD $\beta \to \gamma$ $\mathbf{in}$ F $\mathbf{do}$

      $\quad\quad\mathbf{if}$ $\beta \subseteq Erg$ $\mathbf{then}$ $Erg := Erg \cup \gamma;$

      Ausgabe $\alpha^+ = Erg;$

$\kappa \subseteq X$ ist genau dann ein Superschlüssel, falls gilt:
$\kappa^+ = X$

Attributhülle – Beispiel

- Attributhülle

$$\alpha^+ = \{\, x \mid \text{es gibt } \alpha \to \beta \in F^+ \text{ mit } x \in \beta\}$$

- Beispiel:

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

  – $\text{AttrHülle}(\text{F}, \{\text{B}\})$:

  $Erg = \{B\}$

  Durchlaufe F:

  $A \to C$: $\{B\}$, $B \to A$: $\{A, B\}$, $AB \to C$: $\{A, B, C\}$

  Durchlaufe F nochmal:

  keine Änderung

  Ausgabe $B^+ = \{A, B, C\}$

Attributhülle

$$\alpha^+ = \{\, x \mid \text{es gibt } \alpha \to \beta \in F^+ \text{ mit } x \in \beta\}$$

- Lemma **L1**: Die Attributhülle besitzt folgende Eigenschaft.

  Für jede Teilmenge $V \subseteq \alpha^+$ gilt auch $\alpha \to V \in F^+$.

- Beweis (Skizze):

  Wir beschränken uns auf den Fall $V = \{a_1, a_2\}$. Da $V \subseteq \alpha^+$ ist, gibt es $\beta_1 = \{a_1, \ldots\}$ und $\beta_2 = \{a_2, \ldots\}$ mit $\alpha \to \beta_1,\ \alpha \to \beta_2 \in F^+$. Mit der Reflexivitätsregel (A_1) sind auch $\beta_1 \to \{a_1\}, \beta_2 \to \{a_2\} \in F^+$. Mit der Transitivitätsregel (A_3) und

$$\alpha \to \beta_1, \qquad \beta_1 \to \{a_1\}\alpha \to \beta_2, \qquad \beta_2 \to \{a_2\}$$

  sind auch $\alpha \to \{a_1\}, \alpha \to \{a_2\} \in F^+$. Mit der Vereinigungsregel (A_4) erhalten wir

$$\alpha \to \{a_1, a_2\} \in F^+.$$

Armstrong-Kalkül – Vollständigkeit

- Das Armstrong-Kalkül ist vollständig, d.h. jede von funktionalen Abhängigkeiten F implizierte FD f lässt sich mithilfe des Armstrong-Kalküls ableiten. Es gilt also

$$F \vDash f \Rightarrow F \vdash f$$

- Beweis (durch Kontraposition):
  - Wir zeigen die Aussage $\mathrm{F} \nvdash f \Rightarrow F \nvDash f$. D.h. ist eine funktionale Abhängigkeit f nicht ableitbar, so wird sie auch nicht von F impliziert. Sei dazu $f = \alpha \to \beta$ eine FD mit $F \nvdash f$ und X die Menge aller Attribute aus F und f.
  - Um zu zeigen, dass $F \nvDash f$ gilt, konstruieren wir eine Relation R, in der F gilt, aber f nicht.

  Konstruktion:

  Seien $\alpha^+ = \{a_1, \ldots, a_n\}$ und $X \setminus \alpha^+ = \{b_1, \ldots, \mathrm{b}_m\}$.

  Konstruiere R wie rechts.

R	$a_1, \ldots, a_n$	$b_1, \ldots, b_m$
r	$1, \ldots, 1$	$1, \ldots, 1$
t	$1, \ldots, 1$	$0, \ldots, 0$

Armstrong-Kalkül – Vollständigkeit

- Zeige für Vollständigkeit: $F \nvdash f \Rightarrow F \nvDash f$
- Hilfslemma 1:

  Es gilt $R \vDash F$, d.h. alle FDs in F werden von R erfüllt.

- Beweis (durch Widerspruch):

  Angenommen es existiert $V \to W \in F$ mit $R \nvDash V \to W$. Dann muss für die beiden Tupel $r, t \in R$ gelten, dass $r[V] = t[V]$ und $r[W] \neq t[W]$ ist. Nach Konstruktion von R kann $r[V] = t[V]$ nur gelten, wenn $V \subseteq \{a_1, \dots, a_n\} = \alpha^+$ ist. Ebenso impliziert $r[W] \neq t[W]$, dass $W \nsubseteq \alpha^+$ ist.

  Nach dem vorigen Lemma **L1** folgt aus $V \subseteq \alpha^+$, dass $\alpha \to V \in F^+$ ist. Mit der

  Transivititätsregel (A_3) erhalten wir aus $\alpha \to V,\ V \to W \in F^+$, dass auch $\alpha \to W \in F^+$ ist.

  Dies liefert einen Widerspruch zu $W \nsubseteq \alpha^+$.

Armstrong-Kalkül – Vollständigkeit

- Zeige für Vollständigkeit: $F \nvdash f \Rightarrow F \nvDash f$
- Hilfslemma 2:

  Es gilt $R \nvDash f$, d.h. f wird von R nicht erfüllt.
- Beweis:

  Da f nicht aus F ableitbar ist $(F \nvdash f)$ gilt insbesondere $\beta \nsubseteq \alpha^+$. Also existiert ein $b_k \in \beta$ mit $b_k \in X \setminus \alpha^+$. Direkt aus der Konstruktion von R folgt dann, dass $r[\alpha] = t[\alpha]$ ist, aber $r[b_k] = 1 \neq 0 = t[b_k]$.

  Insgesamt erhalten wir, dass $f = \alpha \to \beta$ von R nicht erfüllt wird.
- Es gilt also $R \vDash F$, aber $R \nvDash f$. Darauf folgt $F \nvDash f$.

Kanonische Überdeckung – Motivation

- Um für zwei Mengen F und G zu entscheiden, ob sie äquivalent sind ($\mathrm{Sat}(F) = \mathrm{Sat}(G)$), reicht es $F^+ = G^+$ zu überprüfen.

Warum?
Freiwillige Übung

- Im Allgemeinen ist die Hülle F^+ einer Menge von FDs sehr groß
- Vor allem bei Datenbankmodifikationen:
  - Überprüfen der Konsistenz anhand von F^+ sehr aufwändig (auch viele triviale Abhängigkeiten)
  - minimale Menge von „erzeugenden“ funktionalen Abhängigkeiten wünschenswert

- Statt der Hülle F^+: kanonische Überdeckung

Kanonische Überdeckung (Definition)

- Sei F eine Menge von funktionalen Abhängigkeiten über einer Attributmenge X. Dann heißt F_c eine *kanonische Überdeckung* von F, falls gilt:
  1. $F_c^+ = F^+$, d.h. $\mathrm{Sat}(F_c) = \mathrm{Sat}(F)$ (äquivalent)
  2. In F_c existieren keine FDs $\alpha \to \beta$ bei denen α oder β überflüssige Attribute enthalten. D.h.
     - Für alle $A \in \alpha$: $\left(F_c \setminus (\alpha \to \beta)\right) \cup (\alpha \setminus A \to \beta) \not\equiv F_c$ (nicht äquivalent)
     - Für alle $B \in \beta$: $\left(F_c \setminus (\alpha \to \beta)\right) \cup (\alpha \to \beta \setminus B) \not\equiv F_c$
  3. Jede linke Seite einer FD ist einzigartig in F_c (sonst ersetze durch Vereinigungen)

- Beispiel:

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

Eine kanonische Überdeckung ist $\mathrm{F_c} = \{A \to C, B \to A\}$.
(Algorithmus: nächste Folie)

Kanonische Überdeckung – Algorithmus

- Eingabe: Menge von FDs F, Ausgabe: Kanonische Überdeckung F_c
  1. Setze $F_c = F$
  2. Führe für jede FD $\alpha \to \beta \in F_c$ eine Linksreduktion durch:
     - Überprüfe für alle $A \in \alpha$, ob A überflüssig ist. D.h. ob gilt: $\beta \subseteq \mathrm{AttrHülle}(F_c, \alpha \setminus A)$

       Falls ja: ersetze $\alpha \to \beta$ durch $\alpha \setminus A \to \beta$.
  3. Führe für jede FD $\alpha \to \beta \in F_c$ eine Rechtsreduktion durch:
     - Überprüfe für alle $B \in \beta$, ob B überflüssig ist. D.h. ob gilt:

       $$B \in \mathrm{AttrHülle}(\big(F_c \setminus (\alpha \to \beta)\big) \cup (\alpha \to \beta \setminus B), \alpha)$$

       Falls ja: ersetze $\alpha \to \beta$ durch $\alpha \to \beta \setminus B$.
  4. Entferne die im 3. Schritt entstandenen FDs der Form $\alpha \to \emptyset$
  5. Fasse über die Vereinigungsregel FDs der Form $\alpha \to \beta_1, \ldots, \alpha \to \beta_n$

     $$\alpha \to \beta_1 \cup \cdots \cup \beta_n$$

Kanonische Überdeckung – Beispiel

- Beispiel:

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

1. Linksreduktion:
   - $A \rightarrow C$: C ist nicht in $(\{A\} \setminus \{A\})^+ = \emptyset^+$. Ok
   - $B \rightarrow A$: Ok
   - $AB \rightarrow C$: Überprüfe A. Ist $C \in (\{A, B\} \setminus \{A\})^+$? Ja, denn

     $$B^+ = \{ \quad \}$$

     Ersetze $AB \rightarrow C$ durch $B \rightarrow C$.

     B muss nun in $B \rightarrow C$ nicht mehr überprüft werden.

- Zwischenergebnis:

$$F_c = \{A \rightarrow C, B \rightarrow A, B \rightarrow C\}$$

Kanonische Überdeckung – Beispiel

- Beispiel:

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

1. Linksreduktion:
   - $A \to C$: C ist nicht in $(\{A\} \setminus \{A\})^+ = \emptyset^+$. Ok
   - $B \to A$: Ok
   - $AB \to C$: Überprüfe A. Ist $C \in (\{A, B\} \setminus \{A\})^+$? Ja, denn

     $$B^+ = \{B, A, C\}$$

     Ersetze $AB \to C$ durch $B \to C$.

     B muss nun in $B \to C$ nicht mehr überprüft werden.

- Zwischenergebnis:

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

Kanonische Überdeckung – Beispiel

- Zwischenergebnis:

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

2. Rechtsreduktion:
   - $A \to C$: Überprüfe C. $\text{AttrHülle}(\{A \to \emptyset, B \to A, B \to C\},\ A)$

     $= \{$

     Also ist C rechts nicht überflüssig.
   - $B \to A$: Analog: A ist rechts nicht überflüssig.
   - $B \to C$: Überprüfe C. $\text{AttrHülle}(\{A \to C, B \to A, B \to \emptyset\},\ B)$

     $= \{$

     Also ist C auf der rechten Seite überflüssig. Ersetze $B \to C$ durch $B \to \emptyset$.

- Neues Zwischenergebnis:

$$F_c = \{A \to C, B \to A, B \to \emptyset\}$$

Kanonische Überdeckung – Beispiel

- Zwischenergebnis:

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

2. Rechtsreduktion:
   - $A \to C$: Überprüfe C. $\mathrm{AttrHülle}(\{A \to \emptyset, B \to A, B \to C\},\ A)$

     $= \{A\}$

     Also ist C rechts nicht überflüssig.
   - $B \to A$: Analog: A ist rechts nicht überflüssig.
   - $B \to C$: Überprüfe C. $\mathrm{AttrHülle}(\{A \to C, B \to A, B \to \emptyset\},\ B)$

     $= \{$

     Also ist C auf der rechten Seite überflüssig. Ersetze $B \to C$ durch $B \to \emptyset$.

- Neues Zwischenergebnis:

$$F_c = \{A \to C, B \to A, B \to \emptyset\}$$

Kanonische Überdeckung – Beispiel

- Zwischenergebnis:

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

2. Rechtsreduktion:
   - $A \to C$: Überprüfe C. $\text{AttrHülle}(\{A \to \emptyset, B \to A, B \to C\},\ A)$

     $= \{A\}$

     Also ist C rechts nicht überflüssig.
   - $B \to A$: Analog: A ist rechts nicht überflüssig.
   - $B \to C$: Überprüfe C. $\text{AttrHülle}(\{A \to C, B \to A, B \to \emptyset\},\ B)$

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

     Also ist C auf der rechten Seite überflüssig. Ersetze $B \to C$ durch $B \to \emptyset$.

- Neues Zwischenergebnis:

$$F_c = \{A \to C, B \to A, B \to \emptyset\}$$

Kanonische Überdeckung – Beispiel

- Zwischenergebnis:

$$F_c = \{A \to C, B \to A, B \to \emptyset\}$$

3. Entferne $\alpha \to \emptyset$:
   - Entferne die Abhängigkeit $B \to \emptyset$
4. Vereinigen:
   - Hier ist nichts zu tun.

- Ergebnis:

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