

Grundlagen von Datenbanken

4. Übung: Algebraische Optimierung



Algebraische Optimierung

Ziel

- Effiziente Ausführung eines algebraischen Ausdrucks
- Minimierung der Größe von Zwischenergebnissen (das Endergebnis soll gleich bleiben!)

Algebraische Optimierung

Ziel

- Effiziente Ausführung eines algebraischen Ausdrucks
- Minimierung der Größe von Zwischenergebnissen (das Endergebnis soll gleich bleiben!)

Voraussetzung

- Abschätzung der Größe von Zwischenergebnissen

Algebraische Optimierung

Ziel

- Effiziente Ausführung eines algebraischen Ausdrucks
- Minimierung der Größe von Zwischenergebnissen (das Endergebnis soll gleich bleiben!)

Voraussetzung

- Abschätzung der Größe von Zwischenergebnissen

Verwendete Daten

- Anzahl der Tupel in einer Relation R : $Card(R)$
- Anzahl der unterschiedlichen Werte eines Attributes A_i : j_i
- Vertauschungsregeln für Operationen (siehe Skript)
z. B.: $\sigma_{P_1}(\sigma_{P_2}(R)) = \sigma_{P_1 \wedge P_2}(R)$

Algebraische Optimierung: Operatorenbaum

$\sigma_{\text{Nachname}=\text{"Müller"}}(\text{Studenten} \bowtie_{\text{Fach}=FID} \text{Fächer})$

Algebraische Optimierung: Operatorenbaum

$$\sigma_{\text{Nachname}=\text{"Müller"}}(\text{Studenten} \bowtie_{\text{Fach}=FID} \text{Fächer})$$

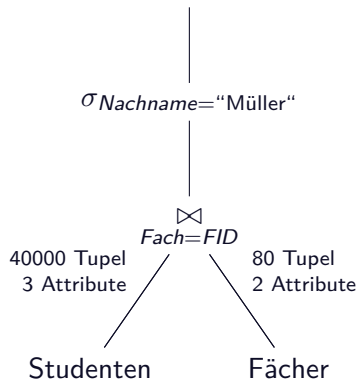
Annahme für das Beispiel

$Card(\text{Studenten}) = 40000$,

$Card(\text{Fächer}) = 80$

Anzahl unterschiedlicher Namen: 250

(bekannt aus dem Data-Dictionary)



Algebraische Optimierung: Operatorenbaum

$\sigma_{\text{Nachname}=\text{"Müller"}}(\text{Studenten} \bowtie_{\text{Fach}=\text{FID}} \text{Fächer})$

Annahme für das Beispiel

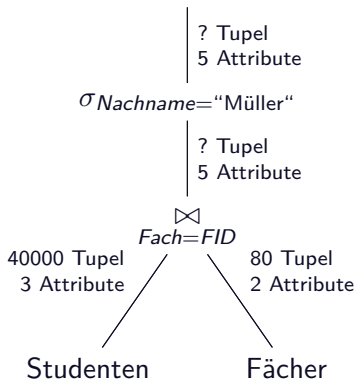
$\text{Card}(\text{Studenten}) = 40000,$

$\text{Card}(\text{Fächer}) = 80$

Anzahl unterschiedlicher Namen: 250
 (bekannt aus dem Data-Dictionary)

Gesucht

Kardinalitäten beliebiger Operationen,
 z. B.: $\text{Card}(\sigma_{\dots}(\text{Studenten} \bowtie \text{Fächer}))$



Selektivitätsfaktor

Motivation

- beschreibt Erwartungswert für die Anzahl der Tupel, die ein Prädikat erfüllen
- basiert auf statistischen Werten
- Annahmen
 - Gleichverteilung der Attributwerte eines Attributes
 - stochastische Unabhängigkeit verschiedener Attribute

Eigenschaften

- $0 \leq SF \leq 1$
- $Card(\sigma_P(R)) = SF(P) \cdot Card(R)$

Berechnung des Selektivitätsfaktors

Prädikate bezüglich eines Attributes

- $SF(A_i = x_i) = \frac{1}{j_i}$, falls Anzahl der Werte j_i für A_i bekannt
- $SF(A_i \geq x_i \wedge A_i \leq x_j) = \frac{x_j - x_i}{\max - \min}$, falls bekannt
- ... (siehe Skript)

Zusammengesetzte Prädikate

- $SF(p(A) \wedge p(B)) = SF(p(A)) \cdot SF(p(B))$
- $SF(p(A) \vee p(B)) = SF(p(A)) + SF(p(B)) - (SF(p(A)) \cdot SF(p(B)))$
- $SF(\neg p(A)) = 1 - SF(p(A))$

Kardinalitätsberechnung beim Verbund

Situation

in der Regel n:1-Verbund zwischen zwei Tabellen:

- TabelleA(PriA, A_1, \dots, A_n , Fremd)
- TabelleB(PriB, B_1, \dots, B_n)
- Referenz: TabelleA.Fremd \rightarrow TabelleB.PriB

Kardinalitätsberechnung beim Verbund

Situation

in der Regel n:1-Verbund zwischen zwei Tabellen:

- $\text{TabelleA}(\underline{\text{PriA}}, A_1, \dots, A_n, \underline{\text{Fremd}})$
- $\text{TabelleB}(\underline{\text{PriB}}, B_1, \dots, B_n)$
- Referenz: $\text{TabelleA.Fremd} \rightarrow \text{TabelleB.PriB}$

Verbund über alle Daten

$$\text{Card}(\text{TabelleA} \underset{\text{Fremd=PriB}}{\bowtie} \text{TabelleB}) = \text{Card}(\text{TabelleA})$$

Kardinalitätsberechnung beim Verbund

Situation

in der Regel n:1-Verbund zwischen zwei Tabellen:

- $\text{TabelleA}(\underline{\text{PriA}}, A_1, \dots, A_n, \underline{\text{Fremd}})$
- $\text{TabelleB}(\underline{\text{PriB}}, B_1, \dots, B_n)$
- Referenz: $\text{TabelleA.Fremd} \rightarrow \text{TabelleB.PriB}$

Verbund über alle Daten

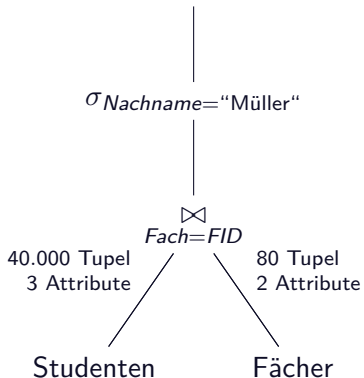
$$\text{Card}(\text{TabelleA} \underset{\text{Fremd=PriB}}{\bowtie} \text{TabelleB}) = \text{Card}(\text{TabelleA})$$

Verbund über eine Teilmenge der Daten

$$\begin{aligned} & \text{Card} \left(\sigma_{P_A}(\text{TabelleA}) \underset{\text{Fremd=PriB}}{\bowtie} \sigma_{P_B}(\text{TabelleB}) \right) \\ &= \text{SF}(P_A) \cdot \text{SF}(P_B) \cdot \text{Card}(\text{TabelleA}) \end{aligned}$$

Kardinalitätsberechnung im Beispiel

$\sigma_{\text{Nachname}=\text{"Müller"}}(\text{Studenten} \bowtie_{\text{Fach}=\text{FID}} \text{Fächer})$

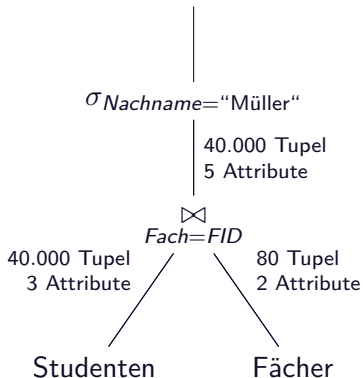


Kardinalitätsberechnung im Beispiel

$\sigma_{\text{Nachname}=\text{"Müller"}}(\text{Studenten} \bowtie_{\text{Fach}=FID} \text{Fächer})$

Berechnung des Verbundes

$Card(\text{Studenten} \bowtie_{\text{Fach}=FID} \text{Fächer})$
 $= Card(\text{Studenten}) = 40.000$



Kardinalitätsberechnung im Beispiel

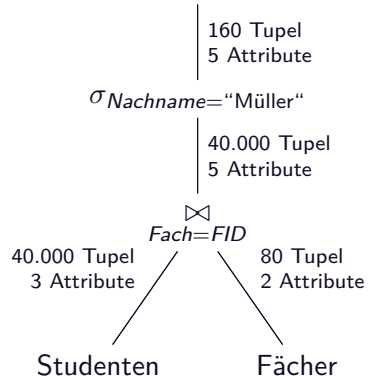
$$\sigma_{\text{Nachname}=\text{"Müller"}}(\text{Studenten} \bowtie_{\text{Fach}=FID} \text{Fächer})$$

Berechnung des Verbundes

$$\begin{aligned} \text{Card}(\text{Studenten} \bowtie_{\text{Fach}=FID} \text{Fächer}) \\ = \text{Card}(\text{Studenten}) = 40.000 \end{aligned}$$

Berechnung der Selektion

$$SF(\text{Nachname} = \text{"Müller"}) = \frac{1}{250}$$



Kardinalitätsberechnung im Beispiel

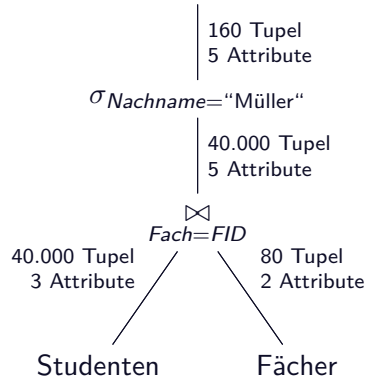
$$\sigma_{\text{Nachname}=\text{"Müller"}}(\text{Studenten} \bowtie_{\text{Fach}=FID} \text{Fächer})$$

Berechnung des Verbundes

$$\begin{aligned} \text{Card}(\text{Studenten} \bowtie_{\text{Fach}=FID} \text{Fächer}) \\ = \text{Card}(\text{Studenten}) = 40.000 \end{aligned}$$

Berechnung der Selektion

$$SF(\text{Nachname} = \text{"Müller"}) = \frac{1}{250}$$



Kardinalitätsberechnung im Beispiel

$$\sigma_{\text{Nachname}=\text{"Müller"}}(\text{Studenten} \bowtie_{\text{Fach}=FID} \text{Fächer})$$

Berechnung des Verbundes

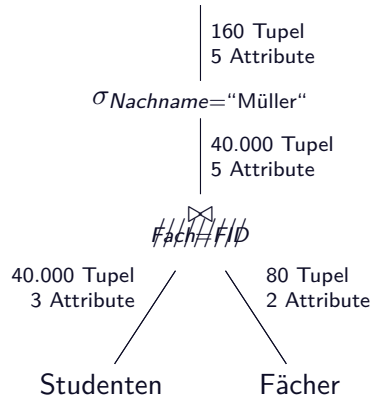
$$\begin{aligned} \text{Card}(\text{Studenten} \bowtie_{\text{Fach}=FID} \text{Fächer}) \\ = \text{Card}(\text{Studenten}) = 40.000 \end{aligned}$$

Berechnung der Selektion

$$SF(\text{Nachname} = \text{"Müller"}) = \frac{1}{250}$$

Spezialfall: natürlicher Join

Join-Attribute sind nur einfach im Ergebnis erhalten



Kardinalitätsberechnung im Beispiel

$$\sigma_{\text{Nachname}=\text{"Müller"}}(\text{Studenten} \bowtie_{\text{Fach}=FID} \text{Fächer})$$

Berechnung des Verbundes

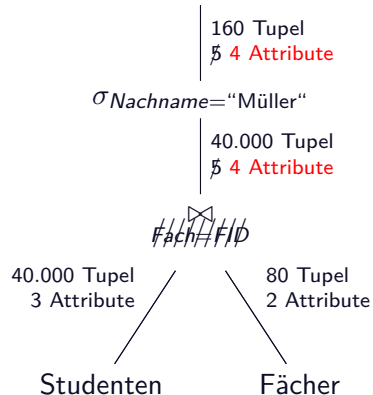
$$\begin{aligned} \text{Card}(\text{Studenten} \bowtie_{\text{Fach}=FID} \text{Fächer}) \\ = \text{Card}(\text{Studenten}) = 40.000 \end{aligned}$$

Berechnung der Selektion

$$SF(\text{Nachname} = \text{"Müller"}) = \frac{1}{250}$$

Spezialfall: natürlicher Join

Join-Attribute sind nur einfach im Ergebnis erhalten



Heuristische Regeln zur Optimierung

- 1 Führe Selektion so früh wie möglich aus
- 2 Führe Projektion so früh wie möglich aus
- 3 (Verknüpfe Folgen von unären Operatoren (soweit möglich))
- 4 Fasse einfache Selektionen auf einer Relation zusammen
- 5 Verknüpfe bestimmte Selektionen mit einem vorausgehenden Kartesischen Produkt zu einem Verbund
- 6 (Berechne gemeinsame Teilbäume nur einmal)
- 7 Bestimme die Verbundreihenfolge so, dass die Anzahl und Größe der Zwischenobjekte minimiert wird
- 8 Verknüpfe bei Mengenoperationen immer zuerst die kleinsten Relationen

Optimierung des Beispiels

$\sigma_{\text{Nachname}=\text{"Müller"}}(\text{Studenten} \bowtie_{\text{Fach}=\text{FID}} \text{Fächer})$

