Grundlagen von Datenbanken

Relationale Algebra und algebraische Optimierung

Relationale Algebra

Überblick

- Selektion: σ
- Projektion: π
- Mengenoperationen: $\cup$, $\cap$, $-$, $\triangleright$, $\div$
- Kartesisches Produkt: $\times$
- Verbund (Join): $\bowtie$
- Umbenennung: ρ

Relationenoperationen: Selektion

Definition

Auswahl von Zeilen einer Relation über ein Prädikat:

$\sigma_P(R) := \{t \in R \mid P(t)\}$

Beispiel

$\sigma_{Wohnort=\text{“Hamburg“} \land Vorname=\text{“Dieter“}}(Studenten)$

$= ?$

Studenten

Matrikel	Vorname	Nachname	Wohnort
28749	Achmed	Barakat	Hamburg
81674	Sarah	Feldbusch	Lübeck
51896	Dieter	Müller	Hamburg

Relationenoperationen: Selektion

Definition

Auswahl von Zeilen einer Relation über ein Prädikat:
$\sigma_P(R) := \{t \in R \mid P(t)\}$

Beispiel

$\sigma_{Wohnort=\text{„Hamburg“} \wedge Vorname=\text{„Dieter“}}(Studenten)$
$= \{(51896, \text{Dieter}, \text{Müller}, \text{Hamburg})\}$

Studenten

Matrikel	Vorname	Nachname	Wohnort
28749	Achmed	Barakat	Hamburg
81674	Sarah	Feldbusch	Lübeck
51896	Dieter	Müller	Hamburg

Relationenoperationen: Projektion

Definition

Auswahl von Spalten einer Relation:

$\pi_{A_1,\ldots,A_k}(R) := \{p \mid \exists t \in R : p = (t[A_1], \ldots, t[A_k])\}$

Beispiel

$\pi_{Wohnort}(Studenten)$

$= ?$

Studenten

Matrikel	Vorname	Nachname	Wohnort
28749	Achmed	Barakat	Hamburg
81674	Sarah	Feldbusch	Lübeck
51896	Dieter	Müller	Hamburg

Relationenoperationen: Projektion

Definition

Auswahl von Spalten einer Relation:

$$\pi_{A_1,\ldots,A_k}(R) := \{p \mid \exists t \in R : p = (t[A_1], \ldots, t[A_k])\}$$

Beispiel

$$\pi_{Wohnort}(Studenten)$$
$$= \{(\text{Hamburg}), (\text{Lübeck})\}$$

Studenten

Matrikel	Vorname	Nachname	Wohnort
28749	Achmed	Barakat	Hamburg
81674	Sarah	Feldbusch	Lübeck
51896	Dieter	Müller	~~Hamburg~~

Relationenoperationen: Mengenoperationen

Definitionen

- $R \cup S := \{t \mid t \in R \vee t \in S\}$ (Vereinigung)
- $R - S := \{t \mid t \in R \wedge t \notin S\}$ (Differenz)
- $R \cap S := \{t \mid t \in R \wedge t \in S\}$ (Durchschnitt)
- $R \triangleright S := (R \cup S) - (R \cap S)$ (Symmetrische Differenz)

Voraussetzung: Vereinigungsverträglichkeit der Relationen!

Beispiel

Stud1 $\cup$ Stud2

Stud1

Matrikel	Vorname
2849	Achmed
8174	Sarah

Stud2

Matrikel	Vorname
8174	Sarah
5196	Dieter

Relationenoperationen: Mengenoperationen

Definitionen

- $R \cup S := \{t \mid t \in R \vee t \in S\}$ (Vereinigung)
- $R - S := \{t \mid t \in R \wedge t \notin S\}$ (Differenz)
- $R \cap S := \{t \mid t \in R \wedge t \in S\}$ (Durchschnitt)
- $R \triangleright S := (R \cup S) - (R \cap S)$ (Symmetrische Differenz)

Voraussetzung: Vereinigungsverträglichkeit der Relationen!

Beispiel

Stud1 $\cup$ Stud2 $=$ {(2849, Achmed), (8174, Sarah), (5196, Dieter)}

Stud1

Matrikel	Vorname
2849	Achmed
8174	Sarah

Stud2

Matrikel	Vorname
8174	Sarah
5196	Dieter

Relationenoperationen: Erweitertes Kartesisches Produkt

Definition

$R \times S := \{k \mid \exists r \in R, s \in S : k = r|s\}$
wobei $r|s := (r_1, \ldots, r_n, s_1, \ldots, s_m)$

Studenten

Matrikel	Vorname	Fach
2849	Achmed	18
8174	Sarah	2

Fächer

FID	Name
18	Informatik
5	Physik

Studenten $\times$ **Fächer**

Matrikel	Vorname	Fach	FID	Name
2849	Achmed	18	18	Informatik
2849	Achmed	18	5	Physik
8174	Sarah	2	18	Informatik
8174	Sarah	2	5	Physik

Relationenoperationen: Verbund (Join)

Definition

$R \underset{A\Theta B}{\bowtie} S = \sigma_{A\Theta B}(R \times S)$
wobei $\Theta \in \{=, \neq, <, >, \geq, \leq\}$

Studenten

Matrikel	Vorname	Fach
2849	Achmed	18
8174	Sarah	2

Fächer

FID	Name
18	Informatik
5	Physik

Studenten $\underset{Fach=FID}{\bowtie}$ **Fächer**

Matrikel	Vorname	Fach	FID	Name
2849	Achmed	18	18	Informatik
2849	Achmed	18	5	Physik
8174	Sarah	2	18	Informatik
8174	Sarah	2	5	Physik

Relationenoperationen: Umbenennung

Definition

Umbenennung der Spalte einer Relation:

$\rho_{B \leftarrow A_i}(R(A_1, \ldots, A_k)) := R(A_1, \ldots, A_{i-1}, B, A_{i+1}, \ldots, A_k)$

Beispiel

$\rho_{Name \leftarrow Vorname}(Studenten)$

Studenten

Matrikel	Vorname	Nachname	Wohnort
28749	Achmed	Barakat	Hamburg
81674	Sarah	Feldbusch	Lübeck
51896	Dieter	Müller	Hamburg

Relationenoperationen: Umbenennung

Definition

Umbenennung der Spalte einer Relation:

$\rho_{B \leftarrow A_i}(R(A_1, \ldots, A_k)) := R(A_1, \ldots, A_{i-1}, B, A_{i+1}, \ldots, A_k)$

Beispiel

$\rho_{Name \leftarrow Vorname}(Studenten)$

Studenten

Matrikel	~~Vorname~~	Nachname	Wohnort
28749	Achmed	Barakat	Hamburg
81674	Sarah	Feldbusch	Lübeck
51896	Dieter	Müller	Hamburg

Relationenoperationen: Umbenennung

Definition

Umbenennung der Spalte einer Relation:

$$\rho_{B \leftarrow A_i}(R(A_1, \ldots, A_k)) := R(A_1, \ldots, A_{i-1}, B, A_{i+1}, \ldots, A_k)$$

Beispiel

$$\rho_{Name \leftarrow Vorname}(Studenten)$$

Studenten

Matrikel	Name	Nachname	Wohnort
28749	Achmed	Barakat	Hamburg
81674	Sarah	Feldbusch	Lübeck
51896	Dieter	Müller	Hamburg

Algebraische Optimierung

Ziel

- Effiziente Ausführung eines algebraischen Ausdrucks (Minimierung der Zwischenergebnisse bei gleichem Endergebnis)

Algebraische Optimierung

Ziel

- Effiziente Ausführung eines algebraischen Ausdrucks (Minimierung der Zwischenergebnisse bei gleichem Endergebnis)

Heuristiken zur Optimierung

I. Führe Selektion so früh wie möglich aus
II. Führe Projektion (ohne Duplikateliminierung) so früh wie möglich aus
III. (Verknüpfe Folgen von unären Operatoren wie Selektion und Projektion)
IV. Fasse einfache Selektionen auf einer Relation zusammen
V. Verknüpfe bestimmte Selektionen mit einem vorausgehenden Kartesischen Produkt zu einem Verbund
VI. (Berechne gemeinsame Teilbäume nur einmal)
VII. Bestimme die Verbundreihenfolge so, dass die Anzahl und Größe der Zwischenobjekte minimiert wird
VIII. Verknüpfe bei Mengenoperationen immer zuerst die kleinsten Relationen

Optimierungsbeispiel (Heuristik I.)

$$\sigma_{Nachname=\text{“Müller“}}(\text{Studenten} \underset{Fach=FID}{\bowtie} \text{Fächer})$$