3.4 Die relationale Algebra

Algebra:

- gegeben eine Menge N ("Anker der Algebra")
- Menge von Operationen $\{\sigma_1,\ldots,\sigma_n\}$ der Form $\sigma_j : N^k \rightarrow N$

Relationale Algebra

- Anker ist die Menge aller Relationen
  $\{(RS, I) \mid I \text{ ist eine Instanz des Schemas } RS\}$
- insgesamt gibt es 6 Grundoperationen

3.4.1 Grundoperationen der relationalen Algebra

- Gegeben seien zwei Relationen R = ($\{A_1,\ldots,A_r\}$, I_R) und S= ($\{B_1,\ldots,B_s\}$, I_S) mit Grad r und Grad s. Es wird durch die folgenden Operationen eine neue Relation T berechnet. Es sei hier angenommen, daß die Attribute des Relationenschema geordnet sind. Dann gilt:

  $I_R = \{(a_1,\ldots,a_r) \mid a_i \in \mathrm{dom}(A_i), 1 \le i \le r\}$ und $I_S = \{ (b_1,\ldots,b_s) \mid b_i \in \mathrm{dom}(B_i), 1 \le i \le s\}$

- **Vereinigung:** $R \cup S$
  - Voraussetzung:
    R und S besitzen gleiches Schema (bis auf Umbenennung der Attribute identisch)
  - $RS_T = RS_S$
  - $I_T = I_R \cup I_S$

- **Differenz:** $R – S$
  - Voraussetzung:
    R und S besitzen gleiches Schema (bis auf Umbenennung der Attribute identisch)
  - $RS_T = RS_S$
  - $I_T = I_R - I_S$

- **Kartesisches Produkt:** $R \times S$
  - $RS_T = RS_R \cup RS_S$
  - $I_T = I_R \times I_S$
- **Projektion:** $\pi_X(\mathbf{R})$
  - Voraussetzung: $X \subseteq RS_R$.
  - $RS_T = X$
  - $I_T = \{r[X] | r \in I_R\}$
  - Einige Anmerkungen zur Schreibweise:
    Da wir angenommen haben, daß das Schema einer Relation geordnet ist, kann die Teilmenge X des Relationenschema RS_R durch Indizes spezifiziert werden.

    $RS_R = \{A_1, A_2, A_3\}$ und $X = \{A_1, A_3\}$: $\pi_{1,\,3}(\mathrm{R}) = \pi_X(\mathrm{R})$

    Die bevorzugte Schreibweise ist jedoch die Menge der Attribute aufzulisten, wobei die Mengenklammern weggelassen werden: $\pi_{A_1,\,A_3}(\mathrm{R}) = \pi_X(\mathrm{R})$
- **Selektion**: $\sigma_{\boldsymbol{F}}(\boldsymbol{R})$

  F bezeichnet eine Formel (F: I_R -> {true,false}), die sich aus Atomen der Form “a op b” zusammensetzt. Dabei sind

– a und b Operanden: Konstanten oder Name eines Attributs

– op ein Vergleichsoperator aus $=, \neq, <, \leq, >, \geq$.

Die Atome können in einer Formel durch

boole’sche Operatoren: $\wedge$, $\vee$, $\neg$

beliebig miteinander verknüpft werden.

Die Ergebnisrelation T ergibt sich dann folgendermaßen:

– $RS_T = RS_R$

– $I_T = \{t | t \in I_R \wedge F(t)\}$

- **Umbenennen** von Attributen $\rho_{B \leftarrow A}(R)$

Attribut A der Relation R wird umbenannt in B. Ansonsten unterscheidet sich das Schema der Relationen nicht.

– Umbenennung unterscheidet sich von den anderen Operatoren dadurch, daß keine neue Instanz erzeugt wird, sondern nur das Schema der Relation verändert wird.

– Operator ist notwendig, wenn eine Relation mehrfach in einer Anfrage vorkommt.

Beispiele

R	A	B	C
	a	b	c
	d	a	f
	c	b	d

S	D	E	F
	b	g	a
	d	a	f

$R \cup S$			
	a	b	c
	d	a	f
	c	b	d
	b	g	a

$R - S$			
	a	b	c
	c	b	d

$R \times S$	A	B	C	D	E	F
	a	b	c	b	g	a
	a	b	c	d	a	f
	d	a	f	b	g	a
	d	a	f	d	a	f
	c	b	d	b	g	a
	c	b	d	d	a	f

$\pi_{A, C}(R)$	A	C
	a	c
	d	f
	c	d

$\sigma_{B = b}(R)$	A	B	C
	a	b	c
	c	b	d

$\rho_{B \leftarrow D}(S)$	B	E	F
	b	g	a
	d	a	f

Beispiele für Anfragen

- Relationenschemata: Städte(<u>SName</u>, SEinw, <u>LName</u>) Länder(<u>LName</u>, LEinw, Partei)

- *Bestimme alle Großstädte und ihre Einwohnerzahlen*

$$\pi_{\text{SName, SEinw}}(\sigma_{\text{SEinw} \geq 500.000}(\text{Städte}))$$

- *In welchem Lande liegt die Stadt Passau?*

$$\pi_{\text{LName}}(\sigma_{\text{SName} = \text{Passau}}(\text{Städte}))$$

- *Bestimme die Namen aller Städte, deren Einwohnerzahl die eines Landes übersteigt.*

$$\pi_{\text{SName}}(\sigma_{\text{SEinw} > \text{LEinw}}(\text{Städte} \times \text{Länder}))$$

- *Finde alle Städtenamen in CDU-regierten Ländern.*

$$\pi_{\text{SName}}(\sigma_{\text{Städte.LName} = \text{Länder.LName}}(\text{Städte} \times \sigma_{\text{Partei} = \text{CDU}}(\text{Länder})))$$

- *Gib alle Städte, die es nur in Hessen gibt?*

$$\pi_{\text{SName}}(\sigma_{\text{LName} = \text{Hessen}}(\text{Städte})) - \pi_{\text{SName}}(\sigma_{\text{LName} \neq \text{Hessen}}(\text{Städte}))$$

3.4.2 Abgeleitete Operationen

Durchschnitt: $R \cap S$

- $R \cap S = R - (R - S)$
- Beispiel:

R	B	C	D
	b	c	a
	b	c	d
	b	f	b
	a	d	c

S	B	C	D
	b	c	d
	b	c	e
	a	d	b

$R \cap S$	B	C	D
	b	c	d

Quotient (Division): $R \div S$

- vereinfachende Annahme:

  r > s, $I_S \neq \emptyset$ und $A_r = B_s$, $A_{r-1} = B_{s-1}, \ldots, A_{r-s+1} = B_1$

- Relationenschema $RS_{R \div S} = \{A_1, \ldots, A_{r-s}\}$
- Instanz des Quotienten:

  $I_{R \div S} := \{(a_1, \ldots, a_{r-s}) \mid \forall (b_1, \ldots, b_s) \in I_S : (a_1, \ldots, a_{r-s}, b_1, \ldots, b_s) \in I_R\}$
- Ableitung des Quotienten durch die Basisoperationen (siehe *Übung*)
- Beispiel:

R	A	B	C	D
	a	b	c	d
	a	b	e	f
	b	c	e	f
	e	d	c	d
	e	d	e	f
	a	b	d	e

S	C	D
	c	d
	e	f

$R \div S$	A	B
	a	b
	e	d

Theta-Join (Verbund): $R_{i\theta j}S$

- Auswahl bestimmter Tupel aus dem kartesischen Produkt $R \times S$:
  $R_{i\theta j}S \; := \; \sigma_{A_i \; \theta \; B_j}(R \times S)$

  mit $\theta \in \{ = , \neq , < , \leq , > , \geq \}$
- Für θ = "=" wird der Join auch als **Equijoin** bezeichnet.
- Beispiel:

R	A	B	C
	1	2	3
	4	5	6
	7	8	9

S	D	E
	3	1
	6	2

$R_{B<D}S$	A	B	C	D	E
	1	2	3	3	1
	1	2	3	6	2
	4	5	6	6	2

Natürlicher Verbund (natural join): $R \bowtie S$

- wichtigste Operation neben der Selektion
- vereinfachende Annahme:

  $A_1 = B_1, \ldots, A_k = B_k$ und $A_j \neq B_i$ für alle j und i mit $k < j \leq r$ und $k < i \leq s$

- Dann ist: $(R \bowtie S) := \pi_{i_{k+1}, i_{k+2}, \ldots, i_{r+s}}\left(\sigma_{R.A_1 = S.B_1 \wedge \ldots \wedge R.A_k = S.B_k}(R \times S)\right)$
- Beispiel:

R	B	C	A
	b	c	a
	b	c	d
	b	f	b
	a	d	c

S	B	C	D
	b	c	d
	b	c	e
	a	d	b

$R \bowtie S$	B	C	A	D
	b	c	a	d
	b	c	a	e
	b	c	d	d
	b	c	d	e
	a	d	c	b

Beispiele

Datenbankschema:

P-M-Zuteilung:

pnr	mnr	Fähigkeit
67	84	3
67	93	2
67	101	3
73	84	5
114	93	5
114	101	3
51	93	2
69	101	2
333	84	3
701	84	2
701	101	2
82	101	2

Abteilungsleiter:

abtnr	pnr
B10	67
A63	333
A64	51

Abteilung:

abtnr	AName
B10	Spielzeug
A63	Computer
A64	Suppen

Personal:

pnr	PName	Vorname	abtnr	Lohn
67	Meier	Helmut	B10	L4
73	Müller	Margot	B10	L5
114	Bayer	Martin	A63	L6
51	Daum	Birgit	A64	L7
69	Störmer	Willi	A64	L6
333	Haar	Hans	A63	L6
701	Reiner	Willi	A64	L6
82	Just	Michael	A64	L6

Maschinen:

mnr	MName
84	Presse
93	Füllanlage
101	Säge

Anfragen

- Gib alle Namen von Personen, die an einer Maschine ausgebildet sind.

$$\pi_{P\text{Name}}(\text{Personal} \bowtie \text{P-M-Zuteilung})$$

- Gib alle Namen der Personen, die an keiner Maschine genügend gut ausgebildet sind.

$$\pi_{P\text{Name}}((\pi_{\text{pnr}}(\text{Personal}) - \pi_{\text{pnr}}(\sigma_{\text{Fähigkeit} < 5}(\text{P-M-Zuteilung}))) \bowtie \text{Personal})$$

- Gib die Namen der Personen aus Abteilung “Suppen”, die an der Maschine mit mnr = 93 ausgebildet sind.

$$\pi_{P\text{Name}}(((\sigma_{\text{AName} = \text{Suppen}}(\text{Abteilung})) \bowtie \text{Personal}) \bowtie \sigma_{\text{mnr} = 93}(\text{P-M-Zuteilung}))$$

- Gib die Namen der Personen, die an der gleichen Maschine ausgebildet sind wie die Person mit pnr = 114.

$$\pi_{P\text{Name}}(\text{Personal} \bowtie ((\pi_{\text{mnr}}(\sigma_{\text{pnr} = 114}(\text{P-M-Zuteilung}))) \bowtie \text{P-M-Zuteilung}))$$

Weitere Join-Operatoren

- bisherige Join-Operatoren werden auch als innere Joins bezeichnet
  - Datensätze ohne Join-Partner gehen verloren

äußere Join-Operatoren (engl.: **outer joins**):

- Datensätze ohne Join-Partner werden (teilweise) berücksichtigt
- linker äußerer Join R ⟕ S: Tupel von R bleiben erhalten
- rechter äußerer Join R ⟖ S: Tupel von S bleiben erhalten
- vollständiger äußerer Join R ⟗ S: Tupel von S und R bleiben erhalten

Semi-Join $R \ltimes S$

$$R \ltimes S = \Pi_{RS_R}(R \bowtie S)$$

- Enthält alle Tupel der Relation R, die an dem Join mit der Relation S beteiligt sind.

Beispiele

R	A	B	C
	a_1	b_1	c_1
	a_2	b_2	c_2

S	C	D	E
	c_1	d_1	e_1
	c_3	d_2	e_2

- linker äußerer Join

R ⟕ S	A	B	C	D	E
	a_1	b_1	c_1	d_1	e_1
	a_2	b_2	c_2	NULL	NULL

- Semi-Join

R ⋉ S	A	B	C
	a_1	b_1	c_1

3.5 Vergleich rel. Algebra und Tupelkalkül

- die relationale Algebra wird als Maß für die Ausdrucksstärke einer Anfragesprache genommen.
- Es gilt: die relationale Algebra, sichere Ausdrücke des Relationenkalkül und sichere Ausdrücke des Bereichskalkül sind äquivalent zueinander
- Sprache L heißt *relational vollständig*, g.d.w. jeder Ausdruck der Relationenalgebra kann in L simuliert werden.
- Sprache L heißt *streng relational vollständig*, g.d.w jeder Ausdruck der relationalen Algebra kann durch einen Ausdruck in der Sprache L simuliert werden.
- Tupel- und Bereichskalkül sind streng relational vollständig

Von relationaler Algebra zum Tupelkalkül

- Seien R_1 und R_2 Relationen mit $R_i = \{t(RS_i) \mid \Psi_i(t)\}, i = 1, 2.$
- Vereinigung:

  $R_1 \cup R_2 = \{t \mid \Psi_1(t) \vee \Psi_2(t)\}$
- Differenz

  $R_1 - R_2 = \{t \mid (\Psi_1(t) \wedge \neg\Psi_2(t))\}$
- Projektion:

  $\pi_{\mathrm{i}_1, \ldots, \mathrm{i}_\mathrm{m}}(R_1) = \{t \mid \exists u(\Psi_1(u) \wedge (t[1] = u[i_1]) \wedge \ \ldots \wedge (t[m] = u[i_m]))\}$
- Selektion:

  $\sigma_{\boldsymbol{F}}(\boldsymbol{R}_1) = \{t \mid (\Psi_1(t) \wedge t[i]\ \theta\ a)\}$ für $F = (t[i]\ \theta\ a)$