Informationssysteme

Sommersemester 2016

Prof. Dr.-Ing. Sebastian Michel
TU Kaiserslautern

smichel@cs.uni-kl.de

Übersicht

- Konjunktive regelbasierte Anfragen:
  $ans(x_{na}) \leftarrow Professoren(x_{pn}, x_{na}, \text{'C4'}, x_{ra})$
- Relationenkalküle: $\{t.name \mid t \in Professoren \wedge t.rang = \text{'C4'}\}$
- **Relationale Algebra**: $\pi_{name}(\sigma_{rang=\text{'C4'}}(Professoren))$
- **SQL:** SELECT name FROM Professoren WHERE rang='C4'

Die Operatoren der relationalen Algebra

- Selektion σ
- Projektion π
- Kreuzprodukt $\times$
- Join (Verbund) $\bowtie$
- Umbenennung ρ
- Differenz $\setminus$
- Division $\div$
- Vereinigung $\cup$
- Schnitt $\cap$
- Semi-Join (linker) $\ltimes$
- Semi-Join (rechter) $\rtimes$
- linker äußerer Join ⟕
- rechter äußerer Join ⟖
- äußerer Join ⟗

Dazu kommen später noch ein paar Erweiterungen.

Grundlagen der relationalen Algebra

Es gibt **Relationen** (Ausprägungen: $R \subseteq D_1 \times D_2 \times ... \times D_n$) und **Relationenschema**:{[attributname1: datentyp1, attributname2: datentyp2, ...]}.

- Das Schema einer Relation wird auch mit $sch(R)$ oder $\mathcal{R}$ beschrieben.
- Relationen sind Mengen (keine Multimengen)
- Reihenfolge der Tupel sowie Reihenfolge der Attribute spielt keine Rolle (Attribute haben Namen).

Die Operatoren und ihre Verwendung:

- Eingabe eines Operators ist eine oder mehrere Relationen.
- **Ausgabe ist auch wieder eine Relation.**
- D.h. Operatoren sind kombinierbar (mit gewissen Regeln).

Beispielrelationen

Professoren			
PersNr	**Name**	**Rang**	**Raum**
2125	Sokrates	C4	226
2126	Russel	C4	232
2127	Kopernikus	C3	310
2133	Popper	C3	52
2134	Augustinus	C3	309
2136	Curie	C4	36
2137	Kant	C4	7

Vorlesungen			
VorlNr	**Titel**	**SWS**	**GelesenVon**
5001	Grundzüge	4	2137
5041	Ethik	4	2125
5043	Erkenntnistheorie	3	2126
5049	Mäeutik	2	2125
4052	Logik	4	2125
5052	Wissenschaftstheorie	3	2126
5216	Bioethik	2	2126
5259	Der Wiener Kreis	2	2133
5022	Glaube und Wissen	2	2134
4630	Die 3 Kritiken	4	2137

Projektion

Projektion

$\pi_{Rang}(Professoren)$

$\pi_{Rang}(Professoren)$
Rang
C4
C3

- Allgemein: $\pi_{A_1,A_2,\ldots,A_k}(R)$, für Attribute A_i aus $\mathcal{R}$
- Wählt aus Attributen der Argumentrelation die Angegebenen aus.

Selektion

Selektion

$\sigma_{Semester>10}(Studenten)$

$\sigma_{Semester>10}(Studenten)$		
MatrNr	Name	Semester
24002	Xenokrates	18
25403	Jonas	12

- Allgemein: σ_F
- Selektionsprädikat F besteht aus
  - Operatoren: $\lor$ (oder) $\land$ (und) $\neg$ (nicht)
  - Arithmetischen Vergleichsoperatoren: $<, \leq, =, >, \geq, \neq$
  - ... und natürlich aus Attributnamen der Argumentrelation oder Konstanten als Operanden

Vereinigung, Schnitt und Mengendifferenz

Zwei Relationen mit gleichem Schema (=Attributnamen und Domänen) können zusammengefasst werden.
Falls Schema nicht gleich: evtl. Attribute umbenennen

Beispiel:

$$\pi_{PersNr,Name}(Assistenten) \cup \pi_{PersNr,Name}(Professoren)$$

=Projektion zweier Relationen auf gemeinsame Attribute, gefolgt von der Vereingung.

Analog dazu: der Schnitt zweier Relationen.

Mengendifferenz:

$$R \setminus S$$

enthält alle Tupel der Relation R, die nicht in Relation S sind.

Kartesisches Produkt (Kreuzprodukt)

Gegeben zwei Relationen R und S:

$$R \times S$$

beinhaltet **ALLE!** $|R| * |S|$ möglichen Paare von Tupeln aus R und S.

Enthält viele (oft auch viele unsinnige) Kombinationen!

Das **resultierende Schema** $sch(R \times S) = sch(R) \cup sch(S) = \mathcal{R} \cup \mathcal{S}$.

Beim Referenzieren der Attribute des resultierenden Schemas wird R.X und S.Y verwendet, insbesondere bei Überlappungen der einzelnen Schemata (=keine Unklarheiten was gemeint ist). Oder Eindeutigkeit durch Umbenennung.

Umbenennung

Gegeben eine Relation R. Dann steht

$$\rho_S(R)$$

für die **Relation R unter neuem Namen S.** Oder: $\rho_{A \leftarrow B}(R)$ für Umbenennung von Attribut B in A.

Umbenennen von Attributed und/oder Relationen ist oftmals notwendig.

Beispiel: Was wird hier berechnet?

$$\pi_{V1.Vorgaenger}(\sigma_{V2.Nachfolger=5216 \wedge V1.Nachfolger=V2.Vorgaenger}$$
$$(\rho_{V1}(voraussetzen) \times \rho_{V2}(voraussetzen)))$$

V1		V2	
Vorgänger	Nachfolger	Vorgänger	Nachfolger

Der natürliche Verbund (Join)

Gegeben zwei Relationen (+ Schemata) mit gemeinsamen Attributen $B_1, \ldots B_k$

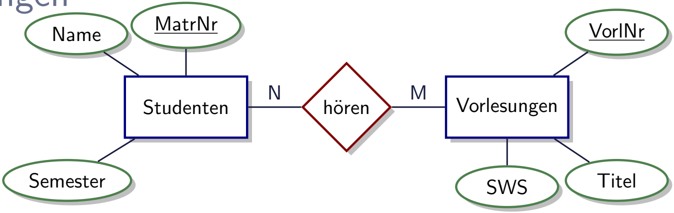
- $R(A_1, ..., A_m, \mathbf{B_1}, ..., \mathbf{B_k})$
- $S(\mathbf{B_1}, ..., \mathbf{B_k}, C_1, ..., C_n)$

Der natürliche Verbund (Join) wird berechnet durch

$$R \bowtie S = \pi_{A_1,...,A_m,R.B_1,...,R.B_k,C_1,...,C_n}(\sigma_{R.B_1=S.B_1 \wedge ... \wedge R.B_k=S.B_k}(R \times S))$$

$R \bowtie S$											
$\mathcal{R} - \mathcal{S}$				$\mathcal{R} \cap \mathcal{S}$				$\mathcal{S} - \mathcal{R}$			
A_1	A_2	...	A_m	B_1	B_2	...	B_k	C_1	C_2	...	C_n
...	...	...	...	...	...	...	...	...	...	...	...

Anwendung von Joins bzgl. Schlüssel/Fremdschlüssel Beziehungen



- **hören:** { MatrNr: integer, VorlNr: integer]}

Im Beispiel hier ist das Attribut MatrNr aus hören ein Fremdschlüssel, der auf MatrNr in der RelatioN Studenten verweise. Analog für Attribut VorlNr, welches auf VorlNr der Relation Vorlesungen verweist.

Mittels Join kann für jedes Tupel aus der Relation hören die zu MatrNr und VorlNr passenden Studenten und Vorlesungen gefunden werden.

Mehrere Joins zusammen

- Es können beliebig viele Relationen durch Join Operatoren verknüpft werden.
- Die Reihenfolge spielt dabei keine Rolle (**Joins sind kommutativ und assoziativ**).

$$(Studenten \bowtie hoeren) \bowtie Vorlesungen$$

$(Studenten \bowtie hoeren) \bowtie Vorlesungen$						
MatrNr	Name	Semester	VorlNr	Titel	SWS	gelesenVon
26120	Fichte	10	5001	Grundzüge	4	2137
27550	Jonas	12	5022	Glaube und Wissen	2	2134
28106	Carnap	3	4052	Wissenschaftstheorie	3	2126
...	...	...	...	...	...	...

Allgemeiner Join (Theta-Join)

Gegeben zwei Relationen (+ Schemata)

- $R(A_1, ..., A_n)$
- $S(B_1, ..., B_m)$

θ ist ein beliebiges Prädikat über den beteiligten Attributen.

$$R \bowtie_\theta S = \sigma_\theta(R \times S)$$

$R \bowtie S$							
$\mathcal{R}$				$\mathcal{S}$			
A_1	A_2	...	A_n	B_1	B_2	...	B_m
...	...	...	...	...	...	...	...

Equi-Join: Join-Prädikat θ darf nur auf Gleichheit (=) prüfen.

Überlegungen zum Join

$$R \bowtie_\theta S = \sigma_\theta(R \times S)$$

Der Join-Operator macht also die relationale Algebra gar nicht ausdrucksstärker, da er durch das Kreuzprodukt mit anschließender Selektion ausgedrückt werden kann, bzw. definiert ist.

Dies gibt auch bereits die erste Idee vor, **wie der Join zweier Relationen berechnet werden kann:**

- Betrachte alle möglichen Paare von Tupeln aus $R \times S$
- Wende Prädikat θ an und schaue ob es “true” liefert.
- Falls ja verknüpfe Tupel und füge resultierendes Tupel in Ergebnis ein.

Weitere Join Typen: Äußere Joins

Diese Join Typen spezifizieren wie mit Tupeln umgegangen wird, die keinen Joinpartner gefunden haben.

- **linker äußerer Join** ($⟕$): auch "partnerlose" Tupel der linken Relation bleiben erhalten
- **rechter äußerer Join** ($⟖$): auch "partnerlose" Tupel der rechten Relation bleiben erhalten
- **vollständiger äußerer Join** ($⟗$): die "partnerlosen" Tupel beider Relationen bleiben erhalten

Natürlicher Join und linker äußerer Join

Natürlicher Join

L

A	B	C
a_1	b_1	c_1
a_2	b_2	c_2

$\bowtie$

R

C	D	E
c_1	d_1	e_1
c_3	d_2	e_2

$=$

Resultat

A	B	C	D	E
a_1	b_1	c_1	d_1	e_1

Linker äußerer Join

L

A	B	C
a_1	b_1	c_1
a_2	b_2	c_2

$⟕$

R

C	D	E
c_1	d_1	e_1
c_3	d_2	e_2

$=$

Resultat

A	B	C	D	E
a_1	b_1	c_1	d_1	e_1
a_2	b_2	c_2	$\perp$	$\perp$

Rechter äußerer und äußerer Join

Rechter äußerer Join

L		
A	B	C
a_1	b_1	c_1
a_2	b_2	c_2

$⟖$

R		
C	D	E
c_1	d_1	e_1
c_3	d_2	e_2

$=$

Resultat				
A	B	C	D	E
a_1	b_1	c_1	d_1	e_1
$\perp$	$\perp$	c_3	d_2	e_2

(Voller) Äußerer Join

L		
A	B	C
a_1	b_1	c_1
a_2	b_2	c_2

$⟗$

R		
C	D	E
c_1	d_1	e_1
c_3	d_2	e_2

$=$

Resultat				
A	B	C	D	E
a_1	b_1	c_1	d_1	e_1
a_2	b_2	c_2	$\perp$	$\perp$
$\perp$	$\perp$	c_3	d_2	e_2

Weitere Join Typen: Semi Joins

Idee: Finde alle Tupel der linken Relation die Joinpartner in der rechten Relation haben.

$$L \ltimes R = \pi_{\mathcal{L}}(L \bowtie R)$$

$L \rtimes R$ ist analog definiert.

Es gilt:

$$L \rtimes R = R \ltimes L$$

allerdings sind Semi-Joins sowie die linken und rechten äußeren Joins **nicht kommutativ!**

Semi-Joins

Semi-Join von L mit R

L		
A	B	C
a_1	b_1	c_1
a_2	b_2	c_2

$\ltimes$

R		
C	D	E
c_1	d_1	e_1
c_3	d_2	e_2

$=$

Resultat		
A	B	C
a_1	b_1	c_1

Semi-Join von R und L

L		
A	B	C
a_1	b_1	c_1
a_2	b_2	c_2

$\rtimes$

R		
C	D	E
c_1	d_1	e_1
c_3	d_2	e_2

$=$

Resultat		
C	D	E
c_1	d_1	e_1

Anti-Joins: Tupel im Ergebnis, falls kein Joinpartner existiert

Anti-Join von L mit R

L

A	B	C
a_1	b_1	c_1
a_2	b_2	c_2

$\triangleright$

R

C	D	E
c_1	d_1	e_1
c_3	d_2	e_2

$=$

Resultat

A	B	C
a_2	b_2	c_2

Anti-Join von R mit L

L

A	B	C
a_1	b_1	c_1
a_2	b_2	c_2

$\triangleleft$

R

C	D	E
c_1	d_1	e_1
c_3	d_2	e_2

$=$

Resultat

C	D	E
c_3	d_2	e_2

$$
\begin{aligned}
e_1 \times e_2 :=& \{y \circ x \mid y \in e_1 \land x \in e_2\} \\
e_1 \bowtie_\theta e_2 :=& \{y \circ x \mid y \in e_1 \land x \in e_2 \land \theta(y,x)\} \\
e_1 \ltimes_\theta e_2 :=& \{y \mid y \in e_1 \land \exists x \in e_2 \; s.t. \; \theta(y,x)\} \\
e_1 \rhd_\theta e_2 :=& \{y \mid y \in e_1 \land \neg\exists x \in e_2 \; s.t. \; \theta(y,x)\} \\
e_1 ⟕_\theta e_2 :=& (e_1 \bowtie e_2) \cup ((e_1 \rhd_\theta e_2) \times \{\perp_{\mathcal{A}(e_2)}\}) \\
e_1 ⟗_\theta e_2 :=& (e_1 \bowtie_\theta e_2) \\
& \cup ((e_1 \rhd_\theta e_2) \times \{\perp_{\mathcal{A}(e_2)}\}) \\
& \cup (\{\perp_{\mathcal{A}(e_1)}\} \times (e_2 \rhd_\theta e_1))
\end{aligned}
$$

$\mathcal{A}(e_i)$ ist hierbei die Menge der Attribute von e_i und $\perp_{\mathcal{A}(e_i)}$ das "NULL" Tupel passend zum Schema von e_i.