Kapitel 2: Grundlagen von Anfragesprachen

Sprachparadigmen

- Relationenalgebra
- Relationenkalkül
- später SQL

2.1 Relationenalgebra

Basisoperatoren

- Attribute aus Relationen herausstreichen: *Projektion*.
- Tupel aus Relationen auswählen: *Selektion*.
- Relationen miteinander verknüpfen: *Verbund*.
- Relationen wie Mengen verarbeiten: *Vereinigung*, *Differenz*.

Projektion

Student

<u>MatrNr</u>	Name	Adresse	Semester
1223	Hans Eifrig	Seeweg 20	2
3434	Lisa Lustig	Bergstraße 11	4
1234	Maria Gut	Am Bächle 1	2

$\Downarrow$

Student'

<u>MatrNr</u>	Name
1223	Hans Eifrig
3434	Lisa Lustig
1234	Maria Gut

Projektion eines Tupels

- Sei $R(X)$ ein Relationsschema, wobei $X = \{A_1, \dots, A_k\}$.
- Sei Y eine Attributmenge, wobei $\emptyset \subset Y \subseteq X$.
- Sei $\mu \in \text{Tup}(X)$ ein Tupel über X.
- Der Ausdruck $\mu[Y]$ heißt *Projektion* des Tupels μ auf Y. Es gilt:

$$\mu[Y] \in \text{Tup}(Y),$$

wobei $\mu[Y](A) = \mu(A), A \in Y$.

Projektion einer Relation

- Sei $r \subseteq \mathrm{Tup}(X)$ eine Relation und $Y \subseteq X$.
- Der Ausdruck $\pi[Y]r$ heißt *Projektion* der Relation r auf Y. Es gilt:

$$\pi[Y]r = \{\mu \in \mathrm{Tup}(Y) \mid \exists \mu' \in r, \text{ so dass } \mu = \mu'[Y]\}.$$

Beispiel

$$r = \begin{array}{ccc} A & B & C \\ \hline a & b & c \\ a & a & c \\ c & b & d \end{array} \qquad \pi[A, C](r) =$$

Selektion

Kurs

KursNr	Institut	Name	Beschreibung
K010	DBIS	Datenbanken	Grundlagen von Datenbanken
K011	DBIS	Informationssysteme	Grundlagen von Informationssystemen
K100	MST	Mikrosystemtechnik	Grundlagen der Mikrosystemtechnik

$\Downarrow$

Kurs'

KursNr	Institut	Name	Beschreibung
K100	MST	Mikrosystemtechnik	Grundlagen der Mikrosystemtechnik

Selektionsbedingung

- Seien $A, B \in X$, $a \in dom(A)$.
- Eine (atomare) *Selektionsbedingung* α (bezüglich X) ist ein Ausdruck der Form $A\,\theta\,B$, bzw. $A\,\theta\,a$, bzw. $a\,\theta\,A$.
- Ein Tupel $\mu \in \mathrm{Tup}(X)$ *erfüllt* eine Selektionsbedingung α, wenn gerade $\mu(A)\,\theta\,\mu(B)$, bzw. $\mu(A)\,\theta\,a$, bzw. $a\,\theta\,\mu(A)$.
- Atomare Selektionsbedingungen können mittels $\wedge$, $\vee$, $\neg$, $($, $)$ zu Formeln verallgemeinert werden.

Beispiel

$$X = \{A, B, C\}.$$
$$\mu_1 = (A \to 2, B \to 2, C \to 1),\ \mu_2 = (A \to 2, B \to 3, C \to 2)$$
$$\alpha_1 = (A = B),\ \alpha_2 = ((B > 1) \wedge (C > 1))$$

Welche Tupel erfüllen welche Selektionsbedingungen?

Selektion

- Sei $r \subseteq \mathsf{Tup}(X)$ eine Relation und α eine Selektionsbedingung zu X.
- Der Ausdruck $\sigma[\alpha]r$ heißt *Selektion* der Relation r bezüglich α. Es gilt:

$$\sigma[\alpha]r = \{\mu \in \mathsf{Tup}(X) \mid \mu \in r \wedge \mu \text{ erfüllt } \alpha\}.$$

Beispiel

$$r = \begin{array}{ccc} A & B & C \\ \hline a & b & c \\ d & a & f \\ c & b & d \end{array}$$

$$\sigma[B = b](r) =$$

Vereinigung und Differenz

- Seien X, Y Attributmengen, wobei $X = Y$ und seien weiter $r \subseteq \text{Tup}(X), s \subseteq \text{Tup}(Y)$ zwei entsprechende Relationen.
- $$r \cup s = \{\mu \in \text{Tup}(X) \mid \mu \in r \vee \mu \in s\}.$$
  $$r - s = \{\mu \in \text{Tup}(X) \mid \mu \in r, \text{ wobei } \mu \notin s\}.$$

Beispiel

$r =$

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

$s =$

A	B	C
b	g	a
d	a	f

$r \cup s =$

$r =$

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

$s =$

A	B	C
b	g	a
d	a	f

$r - s =$

Verbund

Student

MatrNr	Name
1223	Hans Eifrig
3434	Lisa Lustig
1234	Maria Gut

Belegung

MatrNr	KursNr
1223	K010
1234	K010

Kurs

KursNr	Name
K010	Datenbanken
K011	Informations-systeme
K100	Mikrosystem-technik

$\Downarrow$

Student'

MatrNr	Name	KursNr	Name
1223	Hans Eifrig	K010	Datenbanken
1234	Maria Gut	K010	Datenbanken

Verbund

- Seien X, Y Attributmengen; XY sei im Folgenden eine Kurzschreibweise für $X \cup Y$.
- Seien weiter $r \subseteq \text{Tup}(X), s \subseteq \text{Tup}(Y)$ zugehörige Relationen.
- Der *(natürliche) Verbund* $\bowtie$ von r und s ist dann definiert:

$$r \bowtie s = \{\mu \in \text{Tup}(XY) \mid \mu[X] \in r \wedge \mu[Y] \in s\}.$$

Beispiel

$r =$

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

$s =$

C	D
3	1
6	2
4	5

$r \bowtie s =$

Verbund fortgesetzt

Seien X_i, $1 \leq i \leq n$ Formate.

- Sei $X_1 \cap X_2 = \emptyset$

  $r_1 \bowtie r_2 = r_1 \times r_2$.

- $\bowtie_{i=1}^{n} r_i = \{\mu \in \mathrm{Tup}(\cup_{i=1}^{n} X_i) \mid \mu[X_i] \in r_i, 1 \leq i \leq n\}$.

Umbenennung

- Seien $X = \{A_1, \ldots, A_k\}$, $Y = \{B_1, \ldots, B_k\}$ Formate.
- Sei δ eine umkehrbar eindeutige Abbildung von X nach Y, wobei $dom(A) = dom(\delta(A))$. Gilt $\delta(A) = B$, so schreiben wir $A \rightarrow B$.
- Sei $r \subseteq \mathrm{Tup}(X)$ eine Relation zu X.
- Die Umbenennung $\delta[X, Y]$ bezüglich r ist wie folgt:

$$\delta[X,Y]r = \{\mu \in \mathrm{Tup}(Y) \mid \begin{array}{l} \exists \mu' \in r, \text{so dass} \\ \mu'(A_i) = \mu(\delta(A_i)), 1 \leq i \leq k\} \end{array}$$

Beispiel

$X = \{A, B, C\}$, $Y = \{D, E, C\}$ und $\delta = \{A \rightarrow D, B \rightarrow E, C \rightarrow C\}$.

$r =$

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

$\delta[X,Y]r =$

Basisoperatoren

- Selektion, Projektion, Vereinigung, Differenz, Verbund und Umbenennung sind die Basisoperatoren der Relationenalgebra.
- Die Anwendung dieser Operatoren auf Relationen liefert als Ergebnis wiederum eine Relation.
- Die zulässigen Ausdrücke der Relationenalgebra können ausgehend von den Basisoperatoren induktiv definiert werden.
- Wir können andere nützliche Operatoren definiern.

weitere Operatoren

Seien X_i, $1 \leq i \leq n$, Formate und seien $r_i \subseteq \mathrm{Tup}(X_i)$, $1 \leq i \leq n$, Relationen.

- *Durchschnitt*. Sei $X_1 = X_2$.

$$r_1 \cap r_2 =$$

- *θ-Verbund*. Sei $X_1 \cap X_2 = \emptyset$ und sei α eine beliebige Selektionsbedingung über $X_1 \cup X_2$.

$$r \bowtie_\alpha s = \sigma[\alpha](r \times s).$$

Enthält α ausschließlich Gleichheitsvergleiche, dann redet man von einem *Equi-Verbund*.

Division

Seien X_1, X_2 Formate, $X_2 \subset X_1$, $Z = X_1 - X_2$ und weiter $r_2 \neq \emptyset$.

$$\begin{aligned} r_1 \div r_2 &= \{\mu \in \mathrm{Tup}(Z) \mid \{\mu\} \times r_2 \subseteq r_1\} \\ &= \pi[Z]r_1 - \pi[Z](\qquad\qquad - r_1). \end{aligned}$$

Beispiel

$r_1 =$

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	d

$r_2 =$

C	D
c	d
e	f

$r_1 \div r_2 =$

Beispiel

Kurs(<u>KursNr</u>, Institut, Name, Beschreibung)
Belegung(<u>MatrNr</u>, <u>KursNr</u>, Semester, Note)

$\pi[\texttt{MatrNr}](\texttt{Belegung} \div \pi[\texttt{KursNr}]\texttt{Kurs})$

Algebra als Anfragesprache

- Ausdrücke über *Relationsbezeichnern* eines Datenbankschemas.
- Es können nicht alle berechenbaren Transformationen über den Instanzen zweier Datenbankschemata mittels der Algebra ausgedrückt werden. Das bekannteste Beispiel für diese eingeschränkte Mächtigkeit ist das Problem, zu einer beliebigen binären Relation die transitive Hülle zu berechnen.

Äquivalenz

Zwei Ausdrücke der Algebra Q, Q' heißen *äquivalent*, $Q \equiv Q'$, genau dann, wenn für jede Instanz $\mathcal{I}$ der Datenbank gilt:

$$\mathcal{I}(Q) = \mathcal{I}(Q').$$

Beispiele

Sei $\mathrm{attr}(\alpha)$ die in einer Selektionsbedingung α verwendete Menge von Attributen und seien $R, S, T \dots$ Relationsbezeichner mit Formaten X, Y, Z.

- $Z \subseteq Y \subseteq X \Longrightarrow \pi[Z](\pi[Y]R) \equiv$
- $\mathrm{attr}(\alpha) \subseteq Y \subseteq X \Longrightarrow \pi[Y](\sigma[\alpha]R) \equiv$
- $R \bowtie R \equiv$
- $X = Y \Longrightarrow R \cap S \equiv$
- $\mathrm{attr}(\alpha) \subseteq X, \mathrm{attr}(\alpha) \cap Y = \emptyset \Longrightarrow \sigma[\alpha](R \bowtie S) \equiv$

2.2 Relationenkalkül

Syntax

Die Formeln des Relationenkalküls (*R-Formeln*) werden aus Konstanten, Variablen, Relationsbezeichnern, Junktoren $\neg, \wedge, \vee$, Quantoren $\forall, \exists$ und Hilfszeichen '(', ')', ',' gebildet.

- Sei R ein Relationsbezeichner der Stelligkeit k. Seien weiter $a_1, \ldots, a_k$ Konstanten oder Variablen.

  $R(a_1, \ldots, a_k)$ ist eine (atomare) R-Formel.

- Eine *Selektionsbedingung* α ist von der Form $X\theta Y$, oder $X\theta a$, oder $a\theta X$, wobei X, Y Variablen, a eine Konstante und $\theta \in \{=, \neq, \leq, <, \geq, >\}$ ein Vergleichsoperator.

  α ist eine (atomare) R-Formel.

- Sei F eine R-Formel.

  $\neg F$ ist eine R-Formel.

- Sei X eine Variable und F eine R-Formel, die die Variable X enthält, jedoch keinen Ausdruck der Form $\exists X$, bzw. $\forall X$. X heißt in diesem Fall *frei* in der R-Formel F und anderenfalls *gebunden* in F.

  $\exists X\ F$ ist eine ($\exists$-*quantifizierte*) R-Formel.
  $\forall X\ F$ ist eine ($\forall$-*quantifizierte*) R-Formel.

  F ist der *Wirkungsbereich* des $\exists$-, bzw. $\forall$-Quantors.

- Seien F und G R-Formeln und sei $\mathcal{V}_F$, bzw. $\mathcal{V}_G$ die Menge der in F, bzw. G vorhandenen Variablen, wobei die Variablen in $\mathcal{V}_F \cap \mathcal{V}_G$ sowohl in F als auch in G frei.

  Die *Konjunktion* $(F \wedge G)$ ist eine R-Formel.
  Die *Disjunktion* $(F \vee G)$ ist eine R-Formel.

- Für Konjunktion und Disjunktion gilt das Assoziativgesetz; überflüssige Klammerungen lassen wir im Folgenden weg.
- Eine Anfrage Q über einem Datenbank-Schema $\mathcal{R}$ im Relationenkalkül, kurz *Kalkülanfrage*, hat die Form

$$\{(a_1, \ldots, a_n) \mid F\},$$

  wobei F eine R-Formel zu $\mathcal{R}$ und $a_1, \ldots, a_n$ Variablen und Konstanten.
- Die Menge der Variablen unter den a_i muss hierbei gerade die Menge der freien Variablen in F sein.
- Soll ein Format $\{A_1, \ldots, A_n\}$ für die Menge der Antworten definiert werden, so schreiben wir

$$\{(a_1 : A_1, \ldots, a_n : A_n) \mid F\}.$$

Semantik: atomare Kalkülanfragen

- Sei F eine R-Formel mit Variablenmenge $\mathcal{V}_F$. Eine *Variablenbelegung* ν zu F ist eine Funktion über $\mathcal{V}_F$:

$$\nu : \mathcal{V}_F \rightarrow dom.$$

- Wir erweitern ν zusätzlich um die Identität für Konstanten, d.h. für eine beliebige Konstante a gilt $\nu(a) = a$.
- Betrachte *atomare* Kalkülanfragen über einem Relationsschema $R(A_1, \ldots, A_n)$ der Form

$$Q = \{(a_1 \ldots, a_n) \mid R(a_1, \ldots, a_n)\}.$$

Sei r eine Instanz zu R und sei $F = R(a_1, \ldots, a_n)$. Die *Antwort* zu Q bezüglich r, $Q(r)$, ist definiert zu:

$$Q(r) = \{(\nu(a_1), \ldots, \nu(a_n)) \mid \begin{array}{l} \nu \text{ eine Variablenbelegung zu } \mathcal{V}_F \text{ so,} \\ \text{dass } (\nu(a_1), \ldots, \nu(a_n)) \in r\} \end{array}$$

Beispiele

Seien Relationsschemata $R(A, B)$ und $S(B, C)$ gegeben. Seien r, s hierzu betrachtete Instanzen.

- $\pi[A]\sigma[B = 5]R \equiv$
- $\pi[A]R \equiv$
- $\sigma[A = B]R \equiv$
- $R \bowtie S \equiv$
- $R \cup \delta[B \rightarrow A, C \rightarrow B]S \equiv$
- $R - \delta[B \rightarrow A, C \rightarrow B]S \equiv$
- Sei $T(B)$ ein Relationsschema und gelte $t = \pi[B]s$. $R \div T \equiv$

Semantik

- Sei $Q = \{(a_1, \ldots, a_n) \mid F\}$ eine *beliebige* Kalkülanfrage, wobei $a_1, \ldots, a_n$ Variablen und Konstanten.
  Die *Antwort* zu Q bezüglich einer gegebenen Instanz $\mathcal{I}$ ist $Q(\mathcal{I})$ wie folgt:

$$Q(\mathcal{I}) = \{(\nu(a_1), \ldots, \nu(a_n)) \mid \begin{array}{l} \nu \text{ eine Variablenbelegung zu } \mathcal{V}_F \text{ so,} \\ \text{dass } F \text{ unter } \nu \text{ wahr bezüglich } \mathcal{I}\}. \end{array}$$

Beispiel

Seien $R(A, B), S(C, D)$ Relationsschemata mit Instanzen r, s. Sei $Q = \{(X : A, Y : B, V : C, W : D) \mid R(X, Y) \wedge S(V, W) \wedge Y > V\}$ eine Anfrage.

$r =$

A	B
1	2
2	2
2	1

$s =$

C	D
1	1
1	2
3	1

$\stackrel{Q}{\Longrightarrow}$

Wertebereichsunabhängigkeit

- Sei $Q := \{(a_1, \ldots, a_n) \mid F\}$. Sei $\mathcal{I}$ eine Instanz zu $\mathcal{R}$ und *adom* diejenige Menge, die gerade alle Konstanten in Q und alle Konstanten aus $\mathcal{I}$ enthält.
  adom ist der *aktive* Wertebereich von Q; *adom* ist endlich.
- Q, bzw. F, heißen *wertebereichsunabhängig*, wenn für jede beliebige Menge $D \supset adom$ gilt:

$$Q(\mathcal{I}, adom) = Q(\mathcal{I}, D).$$

Beispiel: nicht wertebereichsunabhängige Anfragen

- Sei $R(A)$ ein Relationsschema und Q eine Anfrage der Form
$$\{X \mid \neg R(X)\},$$
wobei $\mathcal{I}(R) = \{1\}$.
- Seien $R(A, B)$ und $S(B, C)$ Relationsschemata. Sei Q eine Anfrage der Form
$$\{(X, Z) \mid \exists Y (R(X, Y) \vee S(Y, Z))\},$$
wobei $\mathcal{I}(R) = \{(1, 1)\}$, bzw. $\mathcal{I}(S) = \emptyset$.

Sicherheit

Ist eine R-Formel F *sicher*, dann ist sie auch wertebereichsunabhängig.

- F enthält keine $\forall$-Quantoren.
- Wenn $F_1 \vee F_2$ Teilformel von F, dann müssen F_1 und F_2 dieselben freien Variablen besitzen.
- Eine Teilformel G einer Formel F heißt *maximal konjunktiv*, wenn F keine konjunktive Teilformel der Form $H \wedge G$ oder $G \wedge H$ enthält.

  Sei $F_1 \wedge \cdots \wedge F_m$, $m \geq 1$, eine maximal konjunktive Teilformel von F. Alle freien Variablen X müssen *begrenzt* sein im folgenden Sinn ($1 \leq j \leq m$):
  - Falls X in einer Formel F_j frei ist, wobei F_j weder ein Vergleichsausdruck noch negiert ist, dann ist X begrenzt.
  - Falls F_j die Form $X = a$ oder $a = X$ hat und a ist eine Konstante, dann ist X begrenzt.
  - Falls F_j die Form $X = Y$ oder $Y = X$ hat und Y ist begrenzt, dann ist auch X begrenzt.

Beispiele

- $\{(X, Y) \mid X = Y \vee R(X, Y)\}$ ist nicht sicher
- $\{(X, Y) \mid X = Y \wedge R(X, Y)\}$ ist sicher
- $\{(X, Y, Z) \mid R(X, Y, Z) \wedge \neg(S(X, Y) \vee T(Y, Z))\}$ ist nicht sicher, jedoch in der äquivalenten Schreibweise

$$R(X, Y, Z) \wedge \qquad \wedge$$

- Die Formel (Division!) $\{X \mid \forall Y(S(Y) \Rightarrow R(X, Y))\}$ ist äquivalent zu

$$\{X \mid \neg\exists Y(\qquad \wedge \qquad)\}$$

Beide Formulierungen sind nicht sicher, jedoch die äquivalente Formulierung

$$\{X \mid \qquad \}$$

ist sicher.