

3.6 Erweiterung der relationalen Algebra

Probleme mit der relationalen Algebra und dem Tupelkalkül

Das bisherige Konzept der relationalen Algebra unterstützt nicht die Formulierung von wichtigen Anfragen.

- Datentyp Relation (Menge von Tupeln) nicht ausreichend.
 - (i) Sortieren der Daten (insbesondere beim Ergebnis einer Anfrage)
 - (ii) Abspeicherung von Duplikaten (die z. B. durch Projektion entstehen)
- Funktionalität der relationalen Algebra nicht ausreichend.
 - (iii) Verdichtung der Daten einer Relation durch Aggregation (Summe, Durchschnitt)

Anmerkung

Wir werden später bei der tatsächlichen Anfragesprache relationaler Systeme (SQL) sehen, daß diese Anforderungen beim Entwurf von SQL berücksichtigt wurden.

Seite 74 von 85

M-Relation

Eine Multi-Relation (M-Relation) R besteht aus einem Relationenschema RS_R und einer Instanz I_R , wobei als Instanz eine Multimenge zulässig ist. Zur Unterscheidung zwischen Mengen und Multimengen werden bei Multimengen spezielle Klammern benutzt: $\langle 1, 2, 1 \rangle$

Sei M eine Multimenge. Dann bezeichnet $V(M, x)$ die Anzahl der Vorkommen eines Elementes x in M .

- Seien M und N Multimengen. Dann gilt $M = N$ genau dann, falls
 $\forall x: V(M, x) = V(N, x)$.

Es wird nun zusätzlich die Operation

- $\delta: MRel \rightarrow Rel$

angeboten. Bei der Transformation wird das Schema der M-Relation direkt an die Relation weitergegeben. Die Instanz der Relation ergibt sich durch Beseitigung von Duplikaten in der Instanz der M-Relation.

Sei $R = (\{A\}, \langle (1), (2), (1) \rangle)$ eine M-Relation und $S = \delta(R)$. Dann ist $S = (\{A\}, \{(1), (2)\})$.

Seite 75 von 85

Erweiterung der relationalen Operatoren

Seien im folgenden R und S zwei M-Relationen und T die M-Relation der Ausgabe.

Vereinigung: $MRel \times MRel \rightarrow MRel$

- Es wird hier wiederum angenommen, daß die Schemata von R und S bis auf Umbennungen identisch sind. Dann gilt $RS_{R \cup S} = RS_S$.
- Die Instanz ergibt sich aus folgender Bedingung:
 $\forall x: V(I_{R \cup S}, x) = V(I_R, x) + V(I_S, x)$.
Bei der Vereinigung werden also Duplikate nicht beseitigt.
- Beispiel: Seien $I_R = \langle (1), (2), (1) \rangle$ und $I_S = \langle (2), (3) \rangle$. Dann ist
 $I_{R \cup S} = \langle (1), (2), (1), (2), (3) \rangle$.

Differenz: $MRel \times MRel \rightarrow MRel$

- Annahme: Schema von R und S sind bis auf Umbennungen identisch.
- $RS_{R-S} = RS_S$
- Die Instanz ergibt sich aus folgender Bedingung:
 $\forall x: (V(I_S, x) = 0 \wedge V(I_{R-S}, x) = V(I_R, x)) \vee (V(I_S, x) > 0 \wedge V(I_{R-S}, x) = 0)$.
Die Differenz entfernt **alle** Elemente aus der ersten Multimenge, die in der zweiten Multimenge vorkommen.

Seite 76 von 85

- Beispiel: $I_R = \langle (1), (2), (1), (2) \rangle$ und $I_S = \langle (2), (3) \rangle$. Dann ist $I_{R-S} = \langle (1), (1) \rangle$.

Kartesisches Produkt: $MRel \times MRel \rightarrow MRel$

- $RS_{R \times S} = RS_R \cup RS_S$.
- $\forall x \forall y: V(I_{R \times S}, xy) = V(I_R, x) \cdot V(I_S, y)$
Es wird analog zu Mengen nun das kartesische Produkt bei Multimengen gebildet.
- Beispiel: $I_R = \langle (1), (2), (1) \rangle$ und $I_S = \langle (2), (3) \rangle$. Dann ist
 $I_{R \times S} = \langle (1, 2), (1, 3), (2, 2), (2, 3), (1, 2), (1, 3) \rangle$.

Projektion: $MRel \rightarrow MRel$

- Das Relationenschema X der Projektion wird explizit beim Operationsaufruf definiert, wobei $X \subseteq RS_R$.
- Es gilt: $\forall x: V(I_{\pi(R)}, x) = V(I_R, x)$. Bei der Projektion wird also jedes Tupel der Eingabe in ein Tupel der Ausgabe überführt.
- Beispiel: Für $I_R = \langle (1, 2), (1, 3) \rangle$ ist $I_{\pi_1(R)} = \langle (1), (1) \rangle$.

Selektion: $MRel \rightarrow MRel$

- $RS_{\sigma_F(R)} = RS_R$

Seite 77 von 85

- Es gilt: $\forall x: (F(x) \wedge V(I_{\sigma_F(R)}, x) = V(I_R, x)) \vee (\neg F(x) \wedge V(I_{\sigma_F(R)}, x) = 0)$. Die Selektion auf M-Relationen entspricht damit der Selektion auf Relationen.
- Beispiel: Sei $RS_R = \{A, B\}$ und $I_R = \langle (1, 2), (1, 3) \rangle$. Dann ist $I_{\sigma_{B=2}(R)} = \langle (1, 2) \rangle$

Umbenennung: MRel $\rightarrow$ MRel

Entspricht der Operationen wie sie für Relationen definiert ist.

Bemerkung

Da eine M-Relation eine Verallgemeinerung der Relation darstellt, sind als Eingabe der Operationen auch Relationen zugelassen. Soll als Ergebnis wiederum dann eine Relation sein, muß aber eine entsprechende Transformation ausgeführt werden.

Beispiel:

Relationenschemata: Städte(SName, SEinw, LName) Länder(LName, LEinw, Partei) .

Finde die Namen der Städte in allen Bundesländern.

$\delta(\pi_{SName}(\sigma_{SEinw \geq 8000}(Städte)))$

Seite 78 von 85

Verallgemeinerung der Projektion

Die Projektion und auch die Umbenennung sind bei M-Relationen Operatoren, die zu jedem Tupel der Eingabe ein Tupel der Ausgabe erzeugen. Solche Abbildungen werden auch als *map* bezeichnet.

Seien R eine M-Relationen, RS_T ein Relationenschema und $f: (RS_R \rightarrow \text{Dom}(R)) \rightarrow (RS_T \rightarrow \text{Dom}(T))$ eine Abbildung. Dann wird durch

$$\mu_f(R)$$

die relationale Map definiert, wobei

- $RS_{\mu_f(R)} = RS_T$
- $\forall x: V(I_{\mu_f(R)}, f(x)) = V(I_R, x)$

Man beachte dabei, daß f eine Funktion ist, die ein Tupel der Relation R auf ein Tupel der Relation T abbildet. Dabei haben wir die Definition des Tupels als eine Abbildung benutzt.

Beispiel

Sei $(\{A, B\}, \{(1, 2), (3, 4), (4, 6)\})$ eine Relation und $\{C, D\}$ ein Schema. Dann wird durch $f(t) = (C \rightarrow t[A] * t[B], D \rightarrow t[A] + t[B])$ eine gewünschte Abbildung definiert. Dadurch wird nun folgende Relation erzeugt: $(\{C, D\}, \{(2, 3), (12, 7), (24, 10)\})$.

Seite 79 von 85

Aggregation

Motivation

Um schnell einen Überblick der Daten einer Relation zu bekommen, soll in einer Anfragesprache auch die Berechnung wichtiger Kennzahlen einer Relation unterstützt werden. Insbesondere im betriebswirtschaftlichen Umfeld sind dabei die Aggregationsoperationen Summe (sum), Durchschnitt (avg), Anzahl (count), Minimum und Maximum von Bedeutung. In der relationalen Algebra sind solche Operationen bisher noch nicht berücksichtigt worden.

Annahmen

Im folgenden betrachten wir nur die Aggregatsoperationen avg, sum, count, die eine numerische Ausgabe besitzen.

Erweiterung der relationalen Algebra

avg, sum, count: MRel $\rightarrow$ numeric

- Bei der Anwendung von sum und avg muß noch zusätzlich ein Attribut aus dem Schema der Relation angegeben werden, auf dem nun das Aggregat gebildet wird.
- Die Operation count liefert die Anzahl der Tupel in der Instanz der Relation.

Seite 80 von 85

Gruppieren

Um nun mehrere Kennzahlen für eine Relation zu berechnen, kann eine Relation in Klassen (Partitionen) aufgeteilt und für jede dieser Partitionen eine Kennzahl berechnet werden.

Eine Partition kann wiederum als Relation angesehen werden, die das Schema von der Quellrelation erbt. Wir verlangen aber, daß die Instanz einer Partition nicht leer ist.

Bei einer Relation R wird eine Partitionierung durch $\{A_1, \dots, A_n\} \subseteq RS_R$ definiert. Eine Partition enthält alle Tupel aus I_R , die bzgl. der Attribute $\{A_1, \dots, A_n\}$ den gleichen Wert besitzen.

Zu einer Aggregatoperation (und einem ausgezeichneten Attribut) wird nun für jede Partition eine Kennzahl berechnet. Diese Kennzahl wird nun zusammen mit den Werten der Partitionierungsattributen in der Ergebnisrelation eingetragen.

Diese Beschreibung skizziert das Verhalten des Operators

- $\gamma_{X, A, agg, B}: \text{MRel} \rightarrow \text{Rel}$

Dabei ist X die Menge der Partitionierungsattribute, A das Aggregationsattribut (optional) und agg die Aggregationsoperation. Die Ausgaberelation besitzt als Schema neben den Attributen aus X noch ein weiteres numerisches Attribut (das dann mit B bezeichnet wird).

Diese Operation läßt sich noch weiter verallgemeinern, so daß für eine Partition gleich mehrere Aggregate berechnet werden können.

Seite 81 von 85

Beispiele:

Wieviel Mitarbeiter gibt es in der Lohngruppe L6?

$$\text{count}(\sigma_{\text{Lohn} = \text{L6}}(\text{Personal}))$$

Wieviel Mitarbeiter gibt es in den einzelnen Abteilungen in der Lohngruppe L6?

$$\gamma_{\text{abtnr}, \text{count}}(\sigma_{\text{Lohn} = \text{L6}}(\text{Personal}))$$

Was ist die durchschnittliche Fähigkeit der Mitarbeiter in einer Abteilung?

S-Relation

Eine S-Relation besteht aus $(R, <_R, [])$, wobei

- R eine M-Relation mit Grad r
- $<_R$ eine Ordnungsrelation auf $D_1 \times \dots \times D_r$
- $[]: \{1, \dots, n\} \rightarrow I_R$ eine bijektive Abbildung mit $n = |I_R|$, so daß $[](i) <_R [](j)$ für $i < j$ gilt. Statt $[](i)$ wird die Schreibweise $I_R[i]$ verwendet.

Durch

- $\omega_{<}: \text{MRel} \rightarrow \text{SRel}$

eine M-Relation R auf eine S-Relation $S = (R, <, [])$ abgebildet, wobei $<$ eine Ordnungsrelation auf R definiert und $[]$ eine durch Sortieren der Relation R erzeugte Abbildung ist.

Beispiel:

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

Seien

- $<_1 = \{((a,b,c), (d,e,f)) \mid a < d\}$
- $<_2 = \{((a,b,c), (d,e,f)) \mid c < f \text{ oder } (c = f) \text{ und } b > e\}$

Ordnungsrelationen. Dann sind

$\omega_{<_1}(R)$	A	B	C
	4	1	3
	7	8	4
	9	2	3

$\omega_{<_2}(R)$	A	B	C
	9	2	3
	4	1	3
	7	8	4

Achtung:

Durch die Operation ω wird eine S-Relation erzeugt. Da aber die Operationen der rel. Algebra nicht für S-Relationen definiert sind, ist es in einem Ausdruck nur möglich, ω als letzte Operation anzuwenden.

Berechne die Liste der Namen aller Städte, deren Einwohnerzahl die eines Landes übersteigt, und sortiere die Liste aufsteigend bzgl. SEinw.

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

Man beachte, daß dieser Ausdruck zuviel an Information liefert, da zusätzlich zu dem Namen der Stadt auch noch der Name des Lands und die Einwohnerzahl zurückgegeben wird.