

Prof. Dr. Guido Moerkotte

Email: moer@pi3.informatik.uni-mannheim.de

Pit Fender

B6, 29, Raum C0.05

68131 Mannheim

Telefon: (0621) 181-2517

Email: pfender@pi3.informatik.uni-mannheim.de

Anfrageoptimierung
Herbst-/Wintersemester 08

3. Übungsblatt
08. Oktober 2008

Aufgabe 1

Gegeben sei folgendes abstraktes relationales Schema:

$R_1(a, c, d, e)$

$R_2(a, b)$

$R_3(b, c)$

$R_4(d)$

$R_5(e)$

Geben Sie für die folgenden Anfragen den Query Graphen an. Klassifizieren Sie den Query Graphen entsprechend der Notation aus der Vorlesung. Wie groß ist jeweils der Suchraum für diese Klasse von Anfragen, wenn ein left-deep bzw. bushy Join-Tree ohne Kreuzprodukte berücksichtigt wird? Existieren effiziente Algorithmen für die Bestimmung der Join-Reihenfolge der Anfrage?

Aufgabe 1 a)

```
SELECT DISTINCT *
FROM R1, R2, R3, R4
WHERE R1.c = R3.c
      AND R2.b = R3.b
      AND R1.d = R4.d;
```

Aufgabe 1 b)

```
SELECT DISTINCT *
FROM R1, R2, R3, R5
WHERE R1.a = R2.a
      AND R1.c = R3.c
      AND R1.e = R5.e
      AND R2.b = R3.b;
```

Aufgabe 1 c)

```

SELECT DISTINCT *
FROM R1, R2, R3, R4, R5
WHERE R1.a = R2.a
      AND R1.c = R3.c
      AND R1.d = R4.d
      AND R1.e = R5.e;

```

Aufgabe 1 d)

Erstellen Sie eine Tabelle, in der sie Join-Ordering-Algorithmen klassifizieren können. D.h. die Tabelle sollte eine Spalte für den Namen des Algorithmus und je eine Spalte für jedes Merkmal eines Join-Ordering-Problems enthalten.

Aufgabe 2

Zeigen Sie, daß die rekursive Formulierung der Kostenfunktion im IKKBZ-Algorithmus korrekt und wohl definiert ist, d.h.

$Cost_H(G) = C_H(G)$ sowie $C_H(S_1S_2)$ und $C_H(S'_1S'_2)$ liefern alle Aufteilung von S in S_1S_2 und $S'_1S'_2$ dieselben Kosten.

Aufgabe 3

Erweitern Sie das Programm vom vorherigen Übungsblatt so, dass wenn möglich Joins erzeugt werden (Klasse HashJoin in tinydb). Messen Sie den Unterschied für die Anfrage

```

select *
from lineitem l,order o
where l.l_orderkey=o.o_orderkey.

```

Aufgabe 4

Gegeben $|R_1|$, $|R_2|$, die Größe des Wertebereichs von $R_1.x$ und $R_2.y$ sowie die Information ob $R_1.x$ bzw. $R_2.y$ Schlüssel von R_1 bzw. R_2 ist.

Aufgabe 4 a)

Wie läßt sich die Selektivität von $\sigma_{R_1.x=c}$ abschätzen, wobei c eine Konstante ist?

Aufgabe 4 b)

Wie läßt sich die Selektivität von $Join_{R_1.x=R_2.y}$ abschätzen?