

Prof. Dr. Bernhard Seeger
 Jochen van den Bercken

Übungen zur Vorlesung
Implementierung von Datenbanksystemen

Abgabe am 5. 4. in der Übung!

Aufgabe 2.1: Kostenschätzung

(3+1+1+2+1+2+2+2+3)

Bestimmen Sie die Größe der Ergebnisrelationen der folgenden Ausdrücke. Nutzen Sie dazu folgende Tabelle:

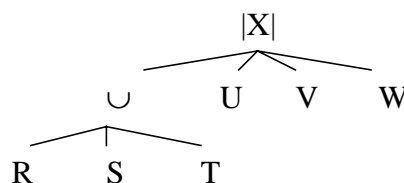
W(a, b)	X(b, c)	Y(c, d)	Z(d, e)
T(W) = 100	T(X) = 200	T(Y) = 300	T(Z) = 400
V(W, a) = 20	V(X, b) = 50	V(Y, c) = 50	V(Z, d) = 40
V(W, b) = 60	V(X, c) = 100	V(Y, d) = 50	V(Z, e) = 100

- $W \bowtie X \bowtie X \bowtie X \bowtie Y \bowtie X \bowtie Z$
- $\sigma_{a=10}(W)$
- $\sigma_{c=20}(Y)$
- $\sigma_{c=20}(Y) \bowtie X \bowtie Z$
- $W \times Y$
- $\sigma_{d>10}(Z)$
- $\sigma_{a=1 \text{ AND } b=2}(W)$
- $\sigma_{a=1 \text{ AND } b>2}(W)$
- $X \bowtie X \bowtie_{X.c < Y.c} Y$

Aufgabe 2.2: Joins

(2+5+2)

- Wie würden Sie die erwartete Größe der Ergebnisrelation bei einem Semijoin schätzen?
- Wenden Sie das Prinzip des dynamischen Programmierens auf Aufgabe 2.1 a) an. Untersuchen Sie einmal nur linkstiefe Bäume, d. h. bei denen die rechte Eingaberelation eines jeden Join-Operators eine Basisrelation ist, und zum anderen alle möglichen Bäume.
- Betrachten Sie folgenden gruppierten Operatorbaum:



Aus wievielen Operatorbäumen kann dieser gruppierte entstanden sein? Beachten Sie, daß die Reihenfolge der Blätter im gruppierten Baum nicht notwendigerweise der Reihenfolge der Blätter im Originalbaum entsprechen muß!

Aufgabe 2.3: Histogramme

(5)

- a) Bestimmen Sie die Größe des Joins $R(a, b) \bowtie S(b, c)$ mit Histogrammen für R.b und S.b. Nehmen Sie $V(R, b) = V(S, b) = 20$ an, und die Histogramme für beide Attribute geben die Häufigkeit der vier am häufigsten vorkommenden Werte an:

	0	1	2	3	4	andere
R.b	5	6	4	5		32
S.b	10	8	5		7	48

Vergleichen Sie diese Schätzung mit der einfachen Schätzung, bei der angenommen werden soll, daß alle 20 Werte gleichhäufig vorkommen, mit $T(R)=52$ und $T(S)=78$?