



Universität Freiburg
Institut für Informatik
Prof. Dr. G. Lausen
Alexander Schätzle
Martin Przyjaciel-Zablocki

Georges-Köhler Allee, Geb. 51
D-79110 Freiburg
lausen@informatik.uni-freiburg.de
schaetzle@informatik.uni-freiburg.de
zablocki@informatik.uni-freiburg.de

Übungen zur Vorlesung
Datenbanken und Informationssysteme
Wintersemester 2013/2014
29.01.2014

13. Aufgabenblatt: Auswertung von Anfrageoperatoren & Transaktionen

Übung 1 (1+2+1 Punkte)

Sei $Sort()$ ein Verfahren zum Sortieren innerhalb des internen Speichers (Primärspeicher). Betrachten Sie unter Verwendung von $Sort()$ das folgende externe Sortierverfahren $extSort()$, mittels dem eine beliebige extern abgelegte Relation sortiert werden kann, die nicht komplett in den Primärspeicher passt.

Seien k Pufferseiten verfügbar und sei N die Anzahl Seiten der Datei, $k < N$.

Durchlauf 0: Es werden jeweils k aufeinander folgende Seiten der Datei gelesen, intern mittels $Sort()$ sortiert und wieder ausgegeben. Dadurch entstehen sortierte Teilfolgen (sog. *Runs*).

Durchlauf $i = 1, 2, \dots$: Verwendet $k - 1$ Puffer-Seiten für die Eingabe und die verbleibende Seite für die Ausgabe. Verschmelze $k - 1$ Runs, die innerhalb des vorherigen Durchlaufs erzeugt wurden, zu einem neuen Run (analog zu *Merge Sort*).

- a) Angenommen es gelte $k = 3$ und eine Seite enthalte maximal 2 Tupel. Demonstrieren Sie das Verfahren für die Eingabe bestehend aus den folgenden $N = 7$ Seiten:

[3, 4], [6, 2], [9, 4], [8, 7], [5, 6], [3, 1], [2]

- b) Wie viele Durchläufe benötigt $extSort()$ zum Sortieren einer Relation mit N Seiten, wenn k Pufferseiten zur Verfügung stehen?
c) Wie viele Seitenzugriffe werden für den Sortiervorgang benötigt?

Übung 2 (1,5+1,5+1,5+1,5 Punkte)

Betrachten Sie den natürlichen Verbund zweier Relationen zu den Schemata $R(A, B)$ und $S(B, C)$. Angenommen ein Tupel zu R benötige 50 Byte und ein Tupel zu S 100 Byte. Die Seitengröße betrage 4 kB, die Relation zu R benötige 10000 Seiten und die Relation zu S 5000 Seiten. Die Blockgröße betrage 100 Seiten und der Datenbankpuffer habe eine Größe von 102 Seiten. Diskutieren Sie für die unterschiedlichen Strategien

- Block-Nested-Loop-Verbund,
- Index-Nested-Loop-Verbund,
- Sort-Merge-Verbund und
- Hash-Verbund

den bei Auswertung des Verbundes entstehenden Aufwand (Anzahl Seitenzugriffe).

Übung 3 (3 Punkte)

Gib zu dem Algebraausdruck $\pi[A](\sigma[A = B](R(A, C) \bowtie (S(B) \bowtie T(C, D))))$ den Anfragebaum an und transformiere diesen schrittweise in einen äquivalenten Anfragebaum, um eine effizientere Auswertung zu erreichen. Diskutieren Sie unterschiedliche Lösungen.

Übung 4 (1+1 Punkte)

Es soll der natürliche Verbund zwischen Relationen zu den Schemata $R(B, C, D)$, $S(C, D, E)$ und $T(D, E, G)$ berechnet werden. Betrachten Sie die unterschiedlichen Varianten der Auswertung $R \bowtie (S \bowtie T)$, $(R \bowtie S) \bowtie T$ und $S \bowtie (R \bowtie T)$.

- Welche dieser Varianten sollte für die Auswertung gewählt werden, wenn keine weiteren Informationen vorhanden sind? Begründen Sie!
- Welche Variante sollte gewählt werden, wenn gilt dass $sel_{RT} < sel_{ST} < sel_{RS}$? Begründen Sie!

Übung 5 (1,5+1,5+2 Punkte)

Betrachten Sie die folgenden Schedules.

S_1 : $R_2Y \ W_2Y \ R_1Y \ W_1Y \ R_2X \ W_2X \ R_1X \ W_1X$.

S_2 : $R_1Y \ W_1Y \ R_1X \ R_2Y \ W_2Y \ W_1X \ R_2X \ W_2X$.

S_3 : $R_1Y \ W_1Y \ R_2Y \ W_2Y \ R_2X \ W_2X \ R_1X \ W_1X$.

- Geben Sie zu jedem der Schedule die Herbrand-Semantik an und zeichnen Sie den Abhängigkeitsgraphen.
- Angenommen, für jede Funktion der Herbrand-Semantik der Form $f(U)$ gilt $f(U) = U + 1$ und für jede Funktion der Form $f(U, V)$ gilt $f(U, V) = (0 * U) + (V + 1)$. Angenommen im Startzustand gilt $X = 0, Y = 0$. Wie lauten die Endzustände?
- Welche der Schedule unter (a) und (b) sind serialisierbar?

Abzugeben durch Einwurf in den Briefkasten Raum 01-025 Gebäude 51 bis spätestens 03.02.2014, 12:00 Uhr.