Institut for Matematik og Datalogi
Syddansk Universitet

October 10, 2011
JFB

Introduction to Computer Science
E11 – Lecture 11

Lecture, October 6, 12:15–14, U26

Rolf Fagerberg covered up through the merging algorithm in section 9.5 in the textbook, including a proof of correctness.

Lecture, October 11, 8:15–10, U26

Rolf Fagerberg will continue lecturing on section 9.5 in the textbook (merging and hashing).

Lecture, November 8, 8:15–10, U26

we will begin on chapter 12 in the textbook.

Discussion section (and study group before that): November 9

The first exercise involves programming. It should be done before you come to discussion section, possibly in your study group. (You may use Python or some other language.)
For the problems below Group 1 should prepare to present Problem 7, on page 428, Problem 3 on page 427, Problem 6 below. Group 2 should prepare to present Problems 54 and 57 on page 436, Problem 8 below. Group 3 should prepare to present Problem 6 on page 428, Problem 1 on page 427, Problem 7 below.

1. Hashing: Write a program to compute the probability of at least one collision when hashing is used with m records and n buckets. (See the calculation on page 427 of your textbook and generalize it.) Assume

that the the hash function spreads data out essentially randomly. Use your program to answer problem 7 on page 428 and problem 57 on page 436. How did you use your program?

2. Hashing: Question 6 on page 428.

3. Hashing: Explain how a poorly chosen hash function can result in a hash storage system becoming little more than a sequential file.

4. Sequential files: Question 3 on page 427 and Problem 54 on page 436.

5. Merging: Question 1 on page 427.

6. Assume sets of numbers are represented by sequential files sorted on element value. For example, the set $\{4, 7, 13, 9, 2\}$ is represented by a sequential file containing $< 2, 4, 7, 9, 13 >$.

   Describe algorithms for constructing $A \cup B$ and $(A \cup B) \cup C$ from A, B and C Note that $(A \cup B) \cup C$ can be done by first computing $A \cup B$ and computing the union of this with C. Instead of giving this solution, process the three files simultaneously, as you do with two files.

7. Assume the database relations A and B each are stored as sequential files of tuples, ordered according to attribute X (which is an attribute of both relations).

   Sketch (details not necessary) an algorithm based on merging for executing the statement

   $$C \leftarrow \text{JOIN } A \text{ and } B \text{ where } A.X = B.X$$

8. Assume again that the database relations A and B each are stored as sequential files, but now no longer ordered on the X attribute.

   Describe an algorithm based on nested loops for executing the statement

   $$C \leftarrow \text{JOIN } A \text{ and } B \text{ where } A.X = B.X$$

   How many comparisons between tuples are performed (as a function of $|A|$ and $|B|$, the numbers of tuples in each relations)?

   Describe how to speed up the algorithm by first using hashing on each relation.

9. Discuss questions 3, 6, 7, and 9 on page 436–437.