

Exercises and hand-ins

Advanced database technology

February 13, 2003

Hand-in

To be handed in at the latest February 20 at 10.00 AM.

Consider the following classic internal sorting algorithm Selection Sort. It works as follows: For $i = 1, \dots, N$ scan the whole input and output the i th smallest element (which, for $i > 1$, is the smallest element larger than the element just output). Some features of this algorithm:

- It does not change the input.
- Uses only memory to store a couple of elements.
- Outputs the sorted list sequentially.

This problem considers the performance of Selection Sort on external memory.

1. Suppose that we have a machine whose internal memory holds only the most recently read block of B elements (plus a few registers). Analyze the I/O complexity of Selection Sort.
2. How does the I/O complexity of Selection Sort change if we have internal memory for M elements, and the virtual memory system always keeps the M/B most recently accessed blocks in internal memory? Argue why your answer is correct.
3. Improve slightly on the above I/O complexity by using another caching strategy for virtual memory. Describe the caching strategy and analyze the I/O complexity when it is used.
4. Devise an External Selection Sort having the abovementioned features (i.e., it should not write anything other than the output to disk, but may make use of all internal memory). It should improve the I/O complexity of the internal version by a factor of M/B .

Other exercises for discussion on February 20

1. GUW 11.3.2
2. GUW 11.4.2
3. GUW 11.4.8
4. Assume $a, b > 1$. Based on the equality $b^{\log_b a} = a$ show the following:
 - (a) $\log_b(a \cdot k) = \log_b a + \log_b k$.
 - (b) $\log_b(a^k) = k \log_b a$.
 - (c) For any $c > 0$ it holds that $\log_b a = \log_c a / \log_c b$.
 - (d) $a^{\log b} = b^{\log a}$.
 - (e) If $c > b$ then $\log_b a > \log_c a$.