CSC6210 (Fall 2009) Assignment 1 (Due 31 Oct)

Problem 1 (40%). Let S be a set of integers. Given an interval $q = [x_1, x_2]$, a *range count* query finds the number of integers in S that fall in q.

Denote by N the number of integers in S, and by B the page size. Give an index structure that consumes $O(N/B)$ space, and answers any range count query in $O(\log_B N)$ I/Os. Describe how to update your structure in $O(\log_B N)$ I/Os per insertion and deletion.

Problem 2 (60%). Let P be a set of N_1 points, and R a set of N_2 rectangles, all in 2d space. We want to find, for each rectangle $r \in R$, the number of points in P covered by r. We also need to report the results nicely by giving a file that contains N_1 records in the form of (r, c), where r is a distinct rectangle in R, and c the answer for r.

Let M be the memory size and B the page size. Give an algorithm to solve the problem in $O(\frac{N_1+N_2}{B} \log_{M/B} \frac{N_1+N_2}{B})$ I/Os.

In case you cannot obtain the above bound, you may still submit your fastest solution, which may earn partial marks depending on how fast it is. The general principle is that (i) the closer it is to the above bound, the higher the mark, and (ii) a solution requiring $\Omega(\frac{N_1}{M}\frac{N_2}{B})$ I/Os earns no mark.

Bonus problem (a correct solution earns 10 extra marks directly in the final exam). Let S be a set of N horizontal segments. As before M is the memory size and B the page size. Give an algorithm to build a persistent B-tree in $O(\frac{N}{B} \log_{M/B} \frac{N}{B})$ I/Os.

Submission requirement. You must typeset your solutions in Latex, and submit a pdf file by email.