August 2009

MADALGO seminar by
Ke Yi, Hong Kong University of Science and Technology

Dynamic indexability and lower bounds for dynamic one-dimensional range query indexes

Abstract:

The B-tree is a fundamental external index structure that is widely used for answering one-dimensional range reporting queries. Given a set of N keys, a range query can be answered in $O(\log_B N/M + K/B)$ I/Os, where B is the disk block size, K the output size, and M the size of the main memory buffer. When keys are inserted or deleted, the B-tree is updated in $O(\log_B N)$ I/Os, if we require the resulting changes to be committed to disk right away. Otherwise, the memory buffer can be used to buffer the recent updates, and changes can be written to disk in batches, which significantly lowers the amortized update cost. A systematic way of batching up updates is to use the logarithmic method, combined with fractional cascading, resulting in a dynamic B-tree that supports insertions in $O(1/B \log N/M)$ I/Os and queries in $O(\log N/M + K/B)$ I/Os. Such bounds have also been matched by several known dynamic B-tree variants in the database literature.
Note that, however, the query cost of these dynamic B-trees is substantially worse than the $O(\log_B N/M + K/B)$ bound of the static B-tree by a factor of $O(\log B)$.

In this paper, we prove that for any dynamic one dimensional range query index structure with query cost $O(q + K/B)$ and amortized insertion cost $O(u/B)$, the tradeoff $q \cdot \log(u/q) = \Omega(\log B)$ must hold if $q = O(\log B)$. For most reasonable values of the parameters, we have $N/M = B^{O(1)}$, in which case our query-insertion tradeoff implies that the bounds mentioned above are already optimal. We also prove a lower bound of $u \cdot \log q = \Omega(\log B)$, which is relevant for larger values of q. Our lower bounds hold in a dynamic version of the indexability model, which is of independent interests. Dynamic indexability is a clean yet powerful model for studying dynamic indexing problems, and can potentially lead to more interesting complexity results.