CSCI5020 External Memory Data Structures: Exercise Spoilers

1 Exercise 1

Problem 1. Do external sort except that, in merging, combine tuples with identical keys. Argue that the cost is as desired.

Problem 2. Apply the algorithm we discussed for $f = \sqrt{M/B}$ recursively.

Problem 3. Argue that there are $n!/((n/k)!)^k$ different results.

2 Exercise 2

Problem 1. Divide $\mathbb{R}^2$ into $\sqrt{M/B}$ slabs. They define $\Theta(M/B)$ multi-slabs, where each multi-slab spans a number of consecutive slabs. Maintain a linked list for each multi-slab during the sweeping process.

Problem 2. First solve the special case where each rectangle $r \in R$ has the form $(-\infty, x] \times (-\infty, y]$.

3 Exercise 3

Problem 1. Persistent B-tree.

Problem 2. Persistent B-tree.

Problem 3. Store something along with each routing element.

Problem 4. How much footprint does each update leave in your structure to the above problem?

4 Exercise 4

Problem 1. First build the base tree and then the secondary structures top-down.

Problem 2. Use a B-tree to index all the data rays by their y-coordinates, and store additional information at each routing element.

Problem 3. Persistent B-tree and filtering search.

Problem 4. Solution to Problem 3 and external interval tree.

5 Exercise 5

Problem 1. Use a B-tree of branching parameter $O((\frac{n}{B})^{1/3} \frac{1}{\log_B n}))$, and then apply the idea of the external range tree.

Problem 2. Generalize the above idea.

Problem 3. First come up with a structure of $O(n/B) + nL/B$ space with query cost $O(\sqrt{nL/B} + kL/B)$. Then think how to recurse.

Problem 4. Top-down.