10M11CI111: Advanced Data Structures

Course Credit: 3
Semester: VII

Introduction

The field of data structures is similar in spirit to algorithms: what's the fastest way to solve problem X. But data structures take a different tact, focusing on ways to organize a typically long-lived corpus of data so that queries can be answered really fast. A typical example are web search engines, which store sophisticated indexes that support a variety of queries (e.g., what documents contain word X) quickly. In contrast to most algorithmic problems, "linear time" is too slow--you can't afford to scan through the entire web to answer a single query--and the goal is typically logarithmic or even constant time per query. Also, space tends to become a more important issue, because the data corpus is usually big, and you don't really want a data structure that is larger than (or even as big as) the data itself.

Course Objectives (Post-conditions)
Knowledge objectives:

1. Student will learn dynamic optimality:(Splay tree, Skip list, tango)
2. Student will learn geometric data structures:(KD tree, Quad Tree, R tree, ray shooting)
3. Student will learn string data structures:(Trie, Compact trie, patricia, suffix tree)
4. Student will learn integer data structures: (Binomial heap, fibonacci heap, soft heap)
5. Student will learn dictionary data structures: (Cuckoo hashing, Bloom filter, Inverted index)
6. Student will learn dynamic graphs:(Link-cut tree, dynamic connectivity)
7. Student will learn succinct data structures: (k-ary tree, multisets, rank/select)
8. Student will learn external memory model:(I/O complexity, Lazy update, B tree, Buffer tree)
9. Student will learn cache-oblivious:(Large matrix multiplication, block tree)

Application objectives:

1. Round-robin MST(Tarzan's implementation using fibonacci heap)
2. Nearest neighbour saerch(approximation and locality sensitive hashing)
3. distributed data structures:(DHT, DBST, *-trees)
4. stream data structures:(Sampling, sketching, fingerprint, wavelet, histogram)

Expected Student Background (Preconditions)

: Basic Data Structures and Computer Programming

Topics Outline:

S NO	Topics	Hrs

1	**Dynamic optimality**: binary search trees, analytic bounds, splay trees, geometric view, greedy algorithm, independent rectangle, Wilber, and Signed Greedy lower bounds; key-independent optimality; O(lg lg *n*)-competitive Tango trees	6
2	**Dictionary data structures**: Cuckoo hashing, Bloom filter, Inverted index	6
3	**String data structures**: suffix tree, suffix array, Inverted index	3
4	I**nteger data structures**: Binomial heap, fibonacci heap, soft heap and their applications to MST and shortest path problems	4
5	**Round-robin MST:** Tarzan's implementation using fibonacci heap and Amortized complexity	2
6	**Succinct data structures:** k-ary tree, multisets, rank/select	2
7	**Geometric data structures**: orthogonal range queries, range trees, interval trees, KD tree, Quad Tree, R tree, ray shooting	6
8	**Nearest neighbour search:** approximation and locality sensitive hashing	2
9	**External memory / cache-oblivious**: I/O complexity, Lazy update, B tree, Buffer tree and cache oblivious implementation of Large matrix multiplication and block tree	4
10	**Distributed data structures:** hash table, trees, stack and lists	5
11	**Stream data structures:** synopsis, sketches, histogram, fingerprint, wavelets, sliding windows etc.	4
	Total	42

References

1. Data Streams: Algorithms and Applications by S. Muthukrishnan†
2. Algorithms and Data Structures for External Memory by Jeffrey Scott Vitter
3. ADVANCED DATA STRUCTURES by PETER BRASS

Evaluation Scheme:

S.No	Examination	Marks
1	T-1	15
2	T-2	25
3	T-3	35
4	*Internal Marks	25

*Internal Marks Breakdown:

Assignments	9 marks (3x3)
Quizzes	12 marks (3x4)
Regularity	4 Marks