Some generl software notes

Nicholas P. Ross

March 21, 2017

Abstract

This is a simple document that discusses the basis and basics of Data Science.

Contents

1 Algorithms

Heap's algorithm
Steinhaus Johnson Trotter algorithm

2 Books

Introduction to Algorithms Paperback, T Cormen
Clean Code: A Handbook of Agile Software Craftsmanship (Robert C. Martin)
The Robert C. Martin Clean Code Collection (Collection) (Robert C. Martin Series)
The Pragmatic Programmer Paperback, by Andrew Hunt

3 What are the top 30 most essential algorithms you must know for competitive programming?

What are the top 30 most essential algorithms you must know for competitive programming?

1. General techniques: greedy algorithms, dynamic programming - dozens of techniques, divide-and-conquer (binary search and other), sorting algorithms - especially merge sort, heap sort and quick sort, partial sort/top-k elements in [expected] linear time.

2. Data structures: arrays, dynamic arrays, stacks, queues, deques, trees, heaps, hash functions and hash tables including rolling hash/polynomial hash for strings/substrings, binary search trees (treap or splay, including the one with implicit keys), segment trees/Fenwick tree, disjoint set union, sqrt-decomposition.

3. Graphs: exploration - DFS/BFS, shortest paths - **Dijkstra**, Bellman-Ford, Floyd-Warshall, spanning trees - **Prims**, Kruskals, flows - Ford-Fulkerson, Edmonds-Karp, min-cost-max-flow, topological sort and strongly connected components, 2-SAT, LCA, bridges and articulation points, eulerian cycle, biconnected graphs - 2-coloring, matching.

4. Strings: KMP, Z-function, polynomial rolling hash, suffix array, suffix tree or suffix automaton.

5. Algebra/Number Theory: Euclids algorithm, fast exponentiation, sieve of Eratosthenes, inverse modulo prime number, solving ax+by=c and alike Diophantine equations, Eulers function computation, Fast Fourier Transform, Gausss algorithm for matrix inversion/solving system of linear equations, chinese theorem about remainders.

6. Geometry: intersecting lines, segments, circles, line and polygon, tangents and common tangents, moving points, lines and segments, sorting by angle, computing lengths and areas, convex hull, scanning line algorithms, fast point inside polygon, finding two closest/farthest points, covering circle.

7. Various: number of combinations, Catalans numbers, inclusion-exclusion formula, Burnsides lemma, Stirling numbers, Grundy numbers for games.

(Also::
Dijkstras
Prims
(B/D)FS
Sieve of Eratshtenos
Binary heaps
Mergesort
“and that’s pretty much it”)

4 References

– 20 articles about core datascience