INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR
CS21003 Algorithms I: End Semester Examination 2022 Spring

Date of Examination: 12^{th} April 2022
Duration: 55 minutes + 5 minutes (for scanning, concatenating, and uploading)
Full Marks: 20
Subject: CS21003 Algorithms I

Part 1

1. (a) Show that the average depth of a node in an n-node binary search tree is $\mathcal{O}(\log n)$.

Solution sketch. Let $\mathcal{D}(n)$ be the average (over all possible binary search trees) sum of depths of all nodes of an n node binary search tree. Any binary search tree with n nodes will have i nodes in the left sub-tree of the root node and $n-i-1$ nodes in the right sub-tree of the root node for some $0 \leqslant i \leqslant n-1$. A crucial observation is that all the n values of i are equally likely on average (do you see why?). Hence, we have the following recurrence.

$$\begin{aligned} \mathcal{D}(n) &= \frac{1}{n}\sum_{i=0}^{n-1}\{\mathcal{D}(i)+\mathcal{D}(n-i-1)\}+n-1 \\ &= \frac{2}{n}\sum_{i=1}^{n-1}\mathcal{D}(i)+n-1 \end{aligned}$$

We can show by substitution method that $\mathcal{D}(n) \leqslant 2n\log n - n$. Since there are n nodes, the average depth of a node in an n-node binary search tree is at most $(2\log n - 1)$. □

(b) We say that a string y has a repetition factor of r is r is the largest integer such that there exists a string x such that y is x concatenated with itself r times. For a pattern $P[1,\dots,m]$, we denote the repetition factor of $P_i = P[1,\dots,i]$ by $\rho(P_i)$. Design a deterministic $\mathcal{O}(m)$ time algorithm to compute $\rho(P_1),\dots,\rho(P_i)$.

[3 + 7 Marks]

Solution sketch. Let r be the repetition factor of $P_i = P[1,\dots,i]$. Then r divides i, $\pi[i] = i - \frac{i}{r}, \rho\left(P_{i-\frac{i}{r}}\right) = r-1$. Hence, the above two conditions are necessary. We will prove that they are sufficient for r to be the repetition factor of P_i. Suppose r is a positive integer which satisfies the above two conditions. We observe that $r=1$ does not satisfy both the equations. Therefore, we have $r \geqslant 2$. Since $\rho\left(P_{i-\frac{i}{r}}\right) = r-1$, we have $P_{i-\frac{i}{r}} = x^{r-1}$ for some string x. Also we have $P[1,\dots,\frac{i}{r}] = x$ since $\rho\left(P_{i-\frac{i}{r}}\right) = r-1$ and $P[1,\dots,\frac{i}{r}] = P[i-\frac{i}{r}+1,\dots,i]$ since $\pi[i] = i-\frac{i}{r}$. Hence, we conclude that $P_i = x^r$. Since π can be computed in $\mathcal{O}(m)$ time, $\rho(P_i)$ for all $i \in [m]$ can be computed in $\mathcal{O}(m)$ time. □

2. (a) Consider the following problem. Input consists of an array $\mathcal{A}[1, \ldots, n]$ of distinct integers and an integer x. If $x = \mathcal{A}[k]$ for some index $k \in [n]$, then output k; otherwise output -1. Design a deterministic $\mathcal{O}(n)$-time algorithm for the above problem.

   The question is wrong. Actual question: If $x = \mathcal{A}[k]$ for some index $k \in [n]$, then find the integer ℓ such that x is the ℓ-th smallest integer.

   Good news: You get full marks in this question.

   Bad news (really? who cares? ☺): I do not need to give solution sketch. Find it yourself if you truly want.

   (b) Show the execution of Knuth-Morris-Pratt algorithm for string matching on text "abaabab" and pattern "abab."

[7 + 3 Marks]

Solution sketch. DIY ☺ □

All the best!!!
