CHENNAI MATHEMATICAL INSTITUTE

M.Sc. / Ph.D. Programme in Computer Science

Entrance Examination, 4 October 2020

Part A has 10 questions of 3 marks each. Part B has 7 questions of 10 marks each. The total marks are 100.
Answers to Part A must be given in the special answer sheet provided for it. Answers to Part B must be written only in the designated space after the corresponding question.

Part A

1. Which of the following languages over the alphabet $\{0, 1\}$ are *not* recognized by deterministic finite state automata (DFA) with *three* states?

   (a) Words which do not have 11 as a contiguous subword

   (b) Binary representations of multiples of three

   (c) Words that have 11 as a suffix

   (d) Words that do not contain 101 as a contiguous subword

2. Consider the following regular expressions over alphabet $\{a, b\}$, where the notation $(a + b)^+$ means $(a + b)(a + b)^*$:

$$r_1 = (a + b)^+ \ a \ (a + b)^*$$
$$r_2 = (a + b)^* \ b \ (a + b)^+$$

   Let L_1 and L_2 be the languages defined by r_1 and r_2, respectively. Which of the following regular expressions define $L_1 \cap L_2$?

   (a) $(a + b)^+ \ a \ (a + b)^* \ b \ (a + b)^+$
   (b) $(a + b)^* \ a \ b \ (a + b)^*$
   (c) $(a + b)^* \ b \ (a + b)^* \ a \ (a + b)^*$
   (d) $(a + b)^* \ a \ (a + b)^* \ b \ (a + b)^*$

3. Some children are given boxes containing sweets. Harish is happy if he gets either gems or toffees. Rekha is happy if she gets both bubble gums and peppermints. Some of the boxes are special, which means that if the box contains either gems or toffees, then it also contains bubble gums and peppermints. If Harish and Rekha are given boxes that are not special, which of the following can we infer?

   (a) Harish is happy

   (b) No bubble gums in Rekha's box

   (c) No toffees in Harish's box

   (d) There are peppermints in Rekha's box

4. In a class, every student likes exactly one novelist and one musician. If two students like the same novelist, they also like the same musician. The class can be divided into novelist groups, each group consisting of all students who like one novelist. Similarly, musician groups can be formed. So each student belongs to one musician group and one novelist group. Which of the following is a valid conclusion?

   (a) There are more musician groups than novelist groups

   (b) There are at least as many novelist groups as musician groups

   (c) For every musician group, there is a bigger novelist group

   (d) For every novelist group, there is a musician group of the same size

5. A boolean function on n variables is a function f that takes an n-tuple of boolean values $x \in \{0,1\}^n$ as input and produces a boolean value $f(x) \in \{0,1\}$ as output.

   We say that a boolean function f is symmetric if, for all inputs $x, y \in \{0,1\}^n$ with the same number of zeros (and hence the same number of ones), $f(x) = f(y)$. What is the number of symmetric boolean functions on n variables?

   (a) $n+1$ (b) $n!$ (c) $\sum_{i=0}^{n} \binom{n}{i}$ (d) 2^{n+1}

6. There are n songs segregated into 3 playlists. Assume that each playlist has at least one song. For all n, the number of ways of choosing three songs consisting of one song from each playlist is:

   (a) $> \frac{n^3}{27}$ (b) $\leq \frac{n^3}{27}$ (c) $\binom{n}{3}$ (d) n^3

The next two questions are based on the following facts.

Basketball shots are classified into *close-range*, *mid-range* and *long-range* shots. Long range shots are worth 3 points, while close-range and mid-range shots are worth 2 points. Of the shots that LeBron James attempts, 45% are close-range, 25% are mid-range, and 30% are long-range. He successfully makes 80% of the close-range shots, 48% of the mid-range shots, and 40% of the long-range shots.

7. What is the probability that a LeBron shot attempt is successful?

   (a) $\frac{1}{2}$ (b) $\frac{4}{5}$ (c) $\frac{3}{5}$ (d) $\frac{4}{7}$

8. What is the probability that a successful 2-point shot attempt by LeBron is a close-range shot?

   (a) $\frac{2}{5}$ (b) $\frac{3}{5}$ (c) $\frac{3}{7}$ (d) $\frac{3}{4}$

9. A fair coin is repeatedly tossed. Each time a head appears, 1 rupee is added to the first bag. Each time a tail appears, 2 rupees are put in the second bag.

   What is the probability that both the bags have the same amount of money after 6 coin tosses?

   (a) $\frac{1}{2^6}$ (b) $\frac{6!}{2! \cdot 4! \cdot 2^6}$ (c) $\frac{2^2}{2^6}$ (d) $\frac{6!}{2^6}$

10. We have a procedure $P(n)$ that makes multiple calls to a procedure $Q(m)$, and runs in polynomial time in n. Unfortunately, a significant flaw was discovered in $Q(m)$, and it had to be replaced by $R(m)$, which runs in exponential time in m. Thankfully, P is still correct when we replace each call to $Q(m)$ with a call to $R(m)$ instead. Which of the following can we *definitely* say about the modified version of P?

    (a) $P(n)$ still runs in polynomial time in n.

    (b) $P(n)$ requires exponential time in n.

    (c) $P(n)$ runs in polynomial time in n if the number of calls made to Q is proportional to $\log n$.

    (d) $P(n)$ runs in polynomial time in n if, for each call $Q(m)$, $m \leq \log n$.

Part B

Answers to Part B must be written only in the designated space after the corresponding question.

1. There are two cities, City X and City Y. Each city has a metro system consisting of three different lines — red line, blue line, and green line. Each station (in both cities) is classified as either *interesting* or *uninteresting*, depending on the places of interest near the station. Both cities have a designated station called *City Centre*. From every station, there is a red line, blue line and a green line, each going to different destinations.

   A sequence of colours represents a journey through the metro system. For example, with the sequence RGGB, one would first take the red line, get down and then take the green line, get down and again take the green line, and so on. Starting from the City Centre, following a sequence of colours will lead to a destination, which may or may not be interesting. Design an algorithm to check if there is a sequence of colours following which one can reach an interesting destination from the City Centre in X, but not from the City Centre in Y.

2. A graph is *finite* if it has a finite number of vertices, and *simple* if it has no self-loops or multiple edges. Assume we are dealing with finite, undirected, simple graphs with at least two vertices. A graph is *connected* if there is a path between any two vertices in the graph, and is *disconnected* otherwise. The *complement* $\overline{G}$ of a graph G has the same vertex set as G, and contains an edge $\{u, v\}$ if and only if G *does not* contain the edge $\{u, v\}$. For the first and third questions below, you will get the credit only if you *also* provide the justification. If your answer is **yes**, drawing an example of such a graph G and its complement $\overline{G}$ suffices as the justification. If your answer is **no**, then you should argue why this is the case.

   (a) Does there exist a graph G with at least two vertices such that *both* G and $\overline{G}$ are *disconnected*? Justify your answer.

   (b) Give an example of a graph G on **four** vertices such that G is isomorphic to its complement $\overline{G}$.

   (c) Does there exist a graph G with at least two vertices such that *both* G and $\overline{G}$ are *connected*? Justify your answer.

3. A graph is *finite* if it has a finite number of vertices, and *simple* if it has no self-loops or multiple edges.

   Prove or disprove: There exists a finite, undirected, simple graph with at least two vertices in which each vertex has a different degree. To give a proof it suffices to draw an example of such a graph. To disprove the result, you should provide an argument as to why such a graph cannot exist.

4. Consider the procedure MYSTERY described in pseudocode below. The procedure takes two non-negative integers as arguments. For a real number x the notation $\lfloor x \rfloor$ denotes the largest integer which is not larger than x.

```
MYSTERY(p, q)
1  if p == 0
2      then return 0
3  r ← ⌊p/2⌋
4  s ← q + q
5  t ← MYSTERY(r, s)
6  if p is even
7      then return t
8      else return t + q
```

   (a) What does MYSTERY(100, 150) return?

   (b) How many times is the statement on line 8 executed in a call to MYSTERY(100, 150)?

   (c) In general, what does MYSTERY(m, n) return for $m, n \geq 0$? Justify your answer with a proof.

5. Let $\Sigma = \{a, b\}$. For two non-empty languages L_1 and L_2 over Σ, we define $Mix(L_1, L_2)$ to be $\{w_1\ u\ w_2\ v\ w_3 \mid u \in L_1, v \in L_2, w_1, w_2, w_3 \in \Sigma^*\}$.

   (a) Give two languages L_1 and L_2 such that $Mix(L_1, L_2) \neq Mix(L_2, L_1)$.

   (b) Show that if L_1 and L_2 are regular, the language $Mix(L_1, L_2)$ is also regular.

   (c) Provide languages L_1 and L_2 that are not regular, for which $Mix(L_1, L_2)$ is regular.

6. A password contains exactly 6 characters. Each character is either a lowercase letter $\{a, b, \ldots, z\}$ or a digit $\{0, 1, \ldots, 9\}$. A *valid* password should contain at least one digit. What is the total number of valid passwords?

   (a) Here is an incorrect answer to the above question. Find the flaw in the argument.

   Let P_i denote the number of passwords where the i-th character is a digit, for $i \in 1, \ldots, 6$.

$$P_1 = 10 \cdot (36)^5$$
$$P_2 = 36 \cdot 10 \cdot (36)^4$$
$$\ldots$$
$$P_6 = (36)^5 \cdot 10$$

   Therefore, total number of valid passwords is $6 \cdot 10 \cdot (36)^5$ (the sum of the right hand sides above).

   (b) What is the correct answer? Provide a justification for your answer. You do not need to simplify your expressions (for example, you can write 26^5, $5!$, etc.).

7. We are given an array of N words $W[1 \cdots N]$, and a length array $L[1 \cdots N]$, where each $L[i]$ denotes the length (number of characters) of $W[i]$. We are also given a *line width* M, which is assumed to be greater than every $L[i]$. We would like to print all these words, in the same order, *without breaking any word across multiple lines.*

To illustrate, suppose $M = 15$, $N = 7$, and the words and lengths are as given below:

Word	`Can`	`you`	`solve`	`this`	`using`	`dynamic`	`programming?`
Length	3	3	5	4	5	7	12

Here are four example ways of laying out this text. Note that there should be a space between adjacent words on the same line.

Layout 1	**Layout 2**	**Layout 3**	**Layout 4**

`Can`
`you`
`solve`
`this`
`using`
`dynamic`
`programming?`
	`Can you`
	`solve this`
	`using dynamic`
	`programming?`
		`Can you solve`
		`this using`
		`dynamic`
		`programming?`
			`Can you solve`
			`this using dynamic`
			`programming?`

Of the four layouts above, Layouts 1, 2 and 3 are valid since each line has at most 15 characters, while Layout 4 is invalid since line 2 requires 18 characters, which will spill over beyond the line width of 15.

Each valid layout has a cost, computed as follows. Let the text be spread out over K lines. For each line i, let s_i be the number of spaces that are appended at the end of the line to make the line exactly M characters long. The cost of the layout is $\sum_{i=1}^{K} s_i^3$. For example, Layout 1 has cost

$$(15-3)^3 + (15-3)^3 + (15-5)^3 + (15-4)^3 + (15-5)^3 + (15-7)^3 + (15-12)^3 = 7326.$$

Layout 2 has cost

$$(15-7)^3 + (15-10)^3 + (15-13)^3 + (15-12)^3 = 672.$$

Layout 3 has cost

$$(15-13)^3 + (15-10)^3 + (15-7)^3 + (15-12)^3 = 672.$$

We can use dynamic programming to compute the smallest cost of any layout for N words $W[1 \cdots N]$ with lengths given by $L[1 \cdots N]$, and line width M. For $1 \le i \le N$, let $C[i]$ denote the minimum cost of any layout for words $W[i \cdots N]$.

(a) Write a recurrence relation for $C[i]$ in terms of $C[i+1], C[i+2], \ldots, C[N]$.

(b) Implement your recurrence as a dynamic programming algorithm.

(c) How much time and space does your algorithm need, in terms of N?