

Technical University of Denmark

Written exam, December 12, 2014.

Course name: Algorithms and data structures.

Course number: 02110.

Aids allow: All written materials are permitted.

Exam duration: 4 hours

Weighting: Question 1: 6% - Question 2: 24% - Question 3: 15% - Question 4: 18% - Question 5: 22% - Question 6: 15%.

The weighting is only an approximative weighting.

You can answer the exam in either Danish or English.

All questions should be answered by filling out the room below the question. As exam paper just hand in this and the following pages filled out. If you need more room you can use extra paper that you hand in together with the exam paper.

Question 1

In the following two questions, there might be more than one statement that is true (there might also be 0).

Question 1.1 Let T be a red-black tree with R red nodes and B black nodes (including the NIL leaves). Which of the following statements are true:

- A $B \leq R$ B $R \leq B$ C $R \geq 1/2B$ D $R \leq 1/2B$ E $B \leq 1/2R$

Question 1.2 Let T be a 2-3-4 tree with n nodes and T_R the corresponding red-black tree. Let R be the number of red nodes in T_R and B the number of black nodes in T_R . Which of the following statements are true:

- A $n = R + B$ B $R \leq n$ C $B \leq n$ D $B \geq n$ E $R = n$

Question 2

Question 2.1 Draw the suffix tree for the string `abcabaca$` (don't replace the labels by indexes into the string, just write the labels on the edges):

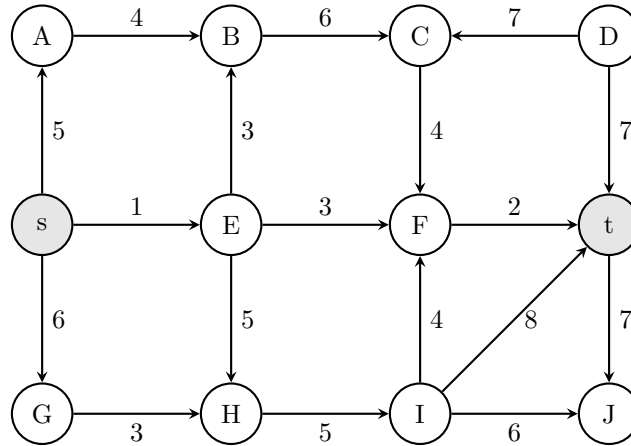
Question 2.2 Compute the prefix function as used in the Knuth-Morris-Pratt algorithm for the string `abcabaca`:

i	a	b	c	a	b	a	c	a
π								

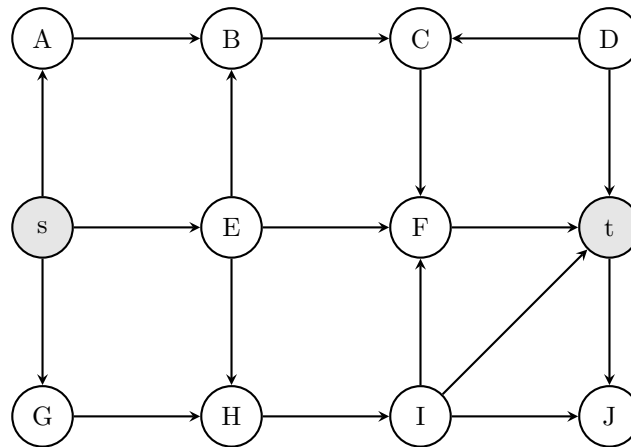
Question 2.3 Draw the string matching finite automata for the string `abcabaca`:

Question 3

Consider the network N below with capacities on the edges.



Question 3.1 Give a maximum flow from s to t in the network N (write the flow for each edge along the edges on the graph below), give the value of the flow, and give a minimum $s - t$ cut (give the partition of the vertices).



value of flow: _____

minimum cut: _____

Question 3.2 Use Edmonds-Karp's algorithm to compute a maximum flow in the network N . For each augmenting path write the nodes on the path and the value you augment the path with in the table below.

augmenting path	value

Question 4

Given two sequences $S = s_1 \cdots s_m$ and $T = t_1 \cdots t_n$, we want to find a string U that contains S and T as subsequences. We call such a string U for a *supersequence* of S and T . A shortest supersequence is a supersequence of minimal length. If $S = AABAB$ and $T = BBABB$ then $U_1 = AABBABB$, $U_2 = ABBABAB$ and $U_3 = AABABBABB$ are all examples of supersequences of S and T . Both U_1 and U_2 are shortest supersequences.

Let $L(i, j)$ denote the length of the shortest supersequence of $s_1 \cdots s_i$ and $t_1 \cdots t_j$.

Question 4.1 Fill out the table below for $L(i, j)$ when $S = CCB$ and $T = BCBC$.

$L(i, j)$	0	1	2	3	4
0					
1					
2					
3					

Question 4.2 Which of the following recurrences correctly computes $L(i, j)$:

$$\text{[A]} \quad L(i, j) = \begin{cases} 0 & \text{if } i = 0 \text{ and } j = 0 \\ L(i - 1, 0) & \text{if } j = 0 \text{ and } i > 0 \\ L(0, j - 1) & \text{if } i = 0 \text{ and } j > 0 \\ L(i - 1, j - 1) + 1 & \text{if } s_i = t_j, i > 0, \text{ and } j > 0 \\ \min\{L(i - 1, j), L(i, j - 1)\} + 1 & \text{if } s_i \neq t_j, i > 0, \text{ and } j > 0 \end{cases}$$

$$\text{[B]} \quad L(i, j) = \begin{cases} i & \text{if } j = 0 \\ j & \text{if } i = 0 \\ L(i - 1, j - 1) & \text{if } s_i = t_j, i > 0, \text{ and } j > 0 \\ \min\{L(i - 1, j), L(i, j - 1)\} + 1 & \text{if } s_i \neq t_j, i > 0, \text{ and } j > 0 \end{cases}$$

$$\text{[C]} \quad L(i, j) = \begin{cases} i & \text{if } j = 0 \\ j & \text{if } i = 0 \\ L(i - 1, j - 1) + 1 & \text{if } s_i = t_j, i > 0, \text{ and } j > 0 \\ \min\{L(i - 1, j), L(i, j - 1)\} + 1 & \text{if } s_i \neq t_j, i > 0, \text{ and } j > 0 \end{cases}$$

Question 4.3 Write the pseudocode for an algorithm based on dynamic programming and the recurrence from Question 4.2 that finds the length of the shortest supersequence of two sequences S and T of length n and m . Analyze the space usage and running time of your algorithm in terms of n and m .

Question 5

A zombie breakout has occurred in the country 1D and you have been asked by the prime minister to help find solutions to stop the infection. You are given a map of the country, with coordinates of all cities. There are X infected cities and Y uninfected cities. The map also contains information about all roads between cities. A road goes directly to from one city to another and all roads are directed. There are R roads. A city is reachable from the capital if you can follow one or more roads from the capital to the city. You can assume that the travel time from the capital to any reachable city is no more than a day.

Question 5.1 The zombies only travel on the roads. The capital is still uninfected and the prime minister wants to know how many roads that must be destroyed to keep the capital uninfected. Give an algorithm to compute the minimum number of roads that has to be destroyed to cut off the capital from the infected cities. Analyze the running time of your algorithm in terms of X , Y , and R . Remember to argue that your algorithm is correct.

Question 5.2 Researchers in the center in the capital have found a vaccine. To get the vaccine to the people in the uninfected cities we need to send 10 doctors to each uninfected city. The doctors can only travel on the roads. They can go through infected cities but at most 50 doctors can go through each infected city per day. Give an algorithm to decide whether we can get the vaccine out to all uninfected cities in one day. Analyze the running time of your algorithm in terms of X , Y , and R . Remember to argue that your algorithm is correct.

Question 6

To find a vaccine against the zombie virus, researchers have been looking at genes from infected and uninfected people. Assume we are given k strings S_1, S_2, \dots, S_k that we know are infected with virus, and one string U that is virus free. All the strings are over an alphabet of size $O(1)$. The total length of the strings is $n = |U| + \sum_{i=1}^k |S_i|$. An *indicator string* V is a string that is a substring of *all* the strings S_1, S_2, \dots, S_k , but *not* a substring of U .

Given the strings S_1, S_2, \dots, S_k, U , and a positive integer ℓ we want to determine whether or not there exists an indicator string V of length *at least* ℓ .

Ex. is $V = BA$ an indicator string for the set of strings below.

	1	2	3	4	5	6	7	8	9	10	11	12
$S_1 =$	A	B	C	A	B	A	C	C	A			
$S_2 =$	C	C	A	B	C	A	A	B	A			
$S_3 =$	B	A	B	C	A	A	B	C	C	A		
$S_4 =$	C	C	A	B	C	A	C	A	B	A		
$S_5 =$	A	B	C	A	A	B	A	C	C	A		
$U =$	A	B	B	C	A	C	C	A	B	C	C	A

Question 6.1 Find an indicator string V of at length at least 3 for the strings above and give the position of its first occurrence in each of the virus infected strings.

$V =$ _____

Question 6.2 Give an algorithm that given the strings S_1, \dots, S_k and U of total length n , and an integer ℓ , finds an indicator string of length at least ℓ or reports that no such indicator string exists.

Analyze the running time of your algorithm. Remember to argue that your algorithm is correct.