

CSE5311 Design and Analysis of Algorithms
Quiz 3
10/29/07

1. The Knuth-Morris-Pratt (KMP) algorithm for String Matching
 - a. Give the algorithm
 - b. Explain the working of the algorithm with the help of an example.
 - c. What is the complexity of the algorithm? Provide proof.

2. The Boyer-Moore Algorithm for String matching
 - a. Give the algorithm
 - b. Explain the working of the algorithm with the help of an example.
 - c. What is the complexity of the algorithm? Provide proof.

3. Let $A = a_1a_2\dots a_n$ and $B = b_1b_2 \dots b_m$ be two strings of characters that come from a finite set (for example, the English alphabet set). The objective is to change A character by character such that it becomes equal to B – called the ‘string-edit’ problem. We allow three types of changes or **edit** steps and assign a cost of ‘1’ to each. The edit steps are: i) *insert* – insert a character into the string; ii) *delete* – delete a character from the string, and iii) *replace* – replace one character with a different character. The number of edits performed is called the ‘edit distance’ between A and B . Provide an efficient algorithm to find the Minimum edit distance between two strings. Briefly describe one application of the ‘Edit Distance’ algorithm.

4. In image processing, ‘segmentation’ is the problem of labeling each pixel as belonging to either the foreground or the background of the scene. The solution to the Segmentation Problem can be obtained by a variation of the Minimum-Cut algorithm used in flow networks. Present an algorithm for Segmentation Problem and analyze it. (Hint: Section 7.10 of the Text book)

5. Suppose we want to replicate a file over a collection of n servers, labeled S_1, S_2, \dots, S_n . To place a copy of the file at server S_i results in a placement cost of c_i , where c_i is an integer greater than 0. Now if a user requests the file from server S_i , and no copy of the file is present at S_i , then servers $S_{i+1}, S_{i+2}, \dots, S_{i+3} \dots$ are searched in order until a copy of the file is finally found, say at server S_j , where $j > i$. This results in an access cost of $j-i$. (Note that the lower-indexed servers $S_{i-1}, S_{i-2} \dots$ are not consulted in this search.) The access cost is 0 if S_i holds a copy of file. We will require that a copy of the file be placed at server S_n , so that all such searches will terminate, at the latest at S_n .
 We’d like to place copies of the files at servers so as to minimize the sum of placement and access costs. Formally, we say that a configuration is a choice, for each server S_i with $i = 1, 2, \dots, n-1$, of whether to place a copy of the file at S_i or not. (Recall that a copy is always placed at S_n .) The total cost of a configuration is the sum of all placement costs for servers with a copy of the file, plus the sum of all access costs associated with all n servers.

Give a polynomial-time algorithm to find a configuration of minimum total cost.

Quiz 3:

You will be assigned one of the above 5 problems.

Task:

1. Solve your problem and submit to the instructor by 11 am Wednesday 10/31/2007. The submission should be preferably PPT form (max 5 slides) by Email to mkumar@uta.edu with subject line 'CSE5311 Quiz 3'. Please copy to the TA as well.
2. Present your solution in the class on 10/31/2007 during the class time. You will be given 10 minutes to make the presentation. To be fair to all, the presenters for 10/31/2007 will be picked using a random method. The remaining presenters will not be allowed to add anything to their slides/submissions after 11 am 10/31/2007.
3. Answer questions by fellow students and instructor.
4. Ask questions.

Grading:

60 points for the solution and correctness,
20 points for clarity in presentation and
20 points for responding to questions and asking questions.

Bonus: In addition to the above,

Implement your algorithm and show code and results. This should also be submitted by 11am 10/31/2007. Your code should work for different input data.

What is the Bonus?

8% of the total points you score in the rest of the course will be added to your score. 8% is the maximum. For example, if your code has a glitch and does not work for different input data, your score will be $y\%$, where $y \leq 8$.

Suppose you score p points (as per the division described earlier) in the entire course, and score $y\%$ the bonus part, your revised score will be $p' = \text{Max} [100, (p + p \times y)]$.

Note: The bonus part will not be added if you do not submit the bonus part by 11 am 10/31/2007, regardless of circumstances.

Problem Assignment to Students

=====

[Bhatty, Fahd Mustafa](#) (5)
[Ganesan Sekar, Vijay Babu](#) (2)
[Jitkajornwanich, Kulsawasd](#) (5)
[Khan, Rosina Surovi](#) (1)
[Lin, Yong](#) (1)
[Meci, Vangjel](#) (2)
[Staton, John Hansford](#) (3)
[Tahir, Samreen](#) (1)
[Varadaraian, Arun](#) (5)
[Venkatachalam, Vijay Shankar](#) (2)
[Wignjosuwito, Agustinus Haris](#) (3)
[Zhang, Dazhi](#) (4)
[Zhou, Lanjiang](#) (4)
