

University of Waterloo
CS240 Winter 2023
Assignment 5 Post Mortem

Question 1 [3+4=7 Marks]

- For part (a), the 2 most common errors were not multiplying the resulting (“rolled”) expression by 100 and not providing explanations for the steps.
- For part (b), most incorrect solutions did not fully explain how to compute the hash values for the string matching aspect, or how the array of suffixes is computed.

Question 2 [4+4=8 Marks]

- This question was generally done well. There were no common errors.

Question 3 [3+6=9 Marks]

- For part (a), the most frequent reason for deducted marks was not providing justification of running time or correctness.
- For part (b), almost all solutions did not include the extra N_{shift} value to maximize the jump. Since this was a non-trivial detail, severe deductions were not given for missing this point.
- For part (b), some answers that were not given full credit did not provide a sufficient explanation of correctness.
- For part (b), many answers that lost marks did not correctly identify all the cases that were possible in a the case of a character mismatch.
- For part (b), other mistakes included updating j before i or comparing i against $m - 1$.

Question 4 [2+4=6 Marks]

- Part (a) was generally done well – the only (infrequent) mistake here was providing a list of strings instead of indices.
- Part (b) was also done well overall. One common misconception that was seen was attempting to sort the suffix array or the list of suffixes. This approach would not have met the required time complexity.

- For part (b), students who lost marks in the justification of correctness did not mention that (i) the suffix arrays is a representation of suffixes in lexicographical order and/or (ii) that if a substring were to occur at least k times, then there are at least k suffixes which have that substring as a prefix.
- For part (b), while many students got a correct algorithm, some missed out on the runtime justification.

Question 5 [3+2+5=10 Marks]

- For part (a), while almost all students got the correct dictionary, some forgot to provide the encoded string as well.
- For part (b), some students only proved that there are 2 leaves at the deepest level without also proving that the 2 leftmost deepest leaves must have the same parent.
- For part (c), most incorrect answers had a runtime complexity of $O(n^2)$ instead of $O(n \log n)$.

Question 6 [3+3=6 Marks]

- For part (a), the most common error here was not encoding the longest possible substring. This was also typically associated with a slightly incorrect dictionary.
- For part (b), many students who got this question wrong did not provide the lexicographically-least string.
- For part (b), some students also did not justify why the compression ratio is the maximum possible (with the crux of the argument being that there are 9 possible substrings of length 2 and all of them occur exactly once in the string).