University of Waterloo
CS240 Spring 2022
Assignment 4 Post-Mortem

This document goes over common errors and general student performance on the assignment questions. We put this together using feedback from the graders once they are done marking. It is meant to be used as a resource to understand what we look at while marking and some common areas where students can improve in.

Question 1 [6 marks]

- This question was generally answered well – many students correctly used properties of compressed tries as part of their justifications, and they also correctly linked the height of the trie to the maximum time it would take to conduct a search operation.

- A few answers did not simplify the search times and/or the height bounds (i.e. the bound was given as $O(\log_{2^t}(n^f))$ instead of $O(\frac{f}{t} \log n)$). No marks were deducted for this, but students are encouraged to try and simplify their answers as much as possible.

- Many students forgot to consider the difference between internal nodes and leaf nodes in a compressed trie. The internal nodes in this question store an array of size $2^t + 1$ while the leaf nodes store the full string, which should give a bound in $O(n2^t + \frac{f}{t} \log n)$. Marks were not deducted for answers of the form $O(n2^t)$.

- Some students did not deduce the number of digits needed to represent every x_i.

- A few students stated that the upper bound for the height is in $O(n)$ – intuitively, this bound should be tighter because the height depends on the number of digits, which in turn depends on the log of the base being used.

- Some students treated f and t as constants and did not include them in their final answers.

Question 2 [2+2+2+2+3+4=15 Marks]

- Parts (a), (b), (c), and (d) were generally answered well. Students are encouraged to put labels next to edges (not on them) since that makes it seem as though the labels are nodes themselves – this makes it more difficult to read.

- For part (d), some students forgot to remove parent nodes that only had one child after deletion. Recall that the internal nodes of compressed tries have at least 2 children.

- For part (e), some students forgot to consider the final end of word character $\$$, which resulted in a final answer of k instead of $k + 1$.

- For part (e), some students forgot to consider the case where $k = 0$ (no marks were deducted for this).

- For part (f), some students stated that $T_{k=0}$, the compressed trie with one leaf, had a height of 0. A compressed trie with 1 leaf has height 1 and is the exception to the rule "a compressed trie with n leaves has $n - 1$ internal nodes". **Please note that the convention for a compressed trie with 1 leaf having a height of 1 is only specific to this assignment. You do not have to worry about such fine distinctions for the final exam, but it is always better to state your assumptions clearly.**

- For part (f), some students got a final answer of $k + 1$ instead of k.

- For part (f), a few students tried to prove the inductive hypothesis instead of stating it and proving the statement for $k + 1$.

Question 3 [3+3+3+3+1=13 Marks]

- For part (a), some students inserted the new element (20) to the tail of the linked list instead of the head; while there was no deduction, students are encouraged to follow the convention in the course notes (which uses the MTF heuristic) and insert new entries into the head of the linked list when using hashing with chaining.

- For part (e), some students forgot to mention that the hash function $\lfloor k/5 \rfloor$ would produce an invalid index 5 when trying to insert 25 into the second hash table.

- For part (e), some incorrect answers included stating that a hash attempt would get stuck in an infinite loop while others made the argument for keys not in the universe. In general, students should only restrict their answers to the keys in the universe.

Question 4 [3+4=7 Marks]

- Part (a) was generally answered very well, with nearly all students correctly identifying that h_1 is a better hash function. Marks were only deducted in cases where insufficient justification had been provided.

- For part (b), some students got an incorrect expression for c_n, which resulted in the final limit being incorrect.

- For part (b), some students summed over the indicator values themselves, instead of equating c_n to the sum of the *expected* indicator values.

- For part (b), a few answers just claimed equality between $\frac{c_n}{n}$ and $\frac{1}{e}$ instead of taking the limit.

Question 5   [4+4+4=12 Marks]

- Part (a) was well answered. Deductions were slightly heavier for wrong answers that only showed one hash table instead of all 7.

- For part (b), many answers that did not receive full marks did not provide sufficient justification, such as stating that there was a finite number of elements in the hash table and that we would eventually end up inserting the largest element in the worst case.

- For part (b), some answers did not use the modified hashing method mentioned in the question and instead structured arguments around regular linear probing.

- For part (c), some students just provided an example without generalizing to arbitrary values of n.

- For part (c), some answers forgot to mention the hash function being used.

- For part (c), many answers that did not receive full marks did not provide a summation or sufficient alternate mathematical justification that proved the $\Omega(n^2)$ bound.