Data Structures and Algorithms – (COMP SCI 2C03)
Winter 2021
Tutorial - 10

April 12, 2021

1. Give a trace for MSD string sort for the keys
   Answer:

Input keys		Keys after sorting characters in the second last place		Sorted keys
no		al		ai
is		ai		al
th		co		co
ti		fo		fo
fo		go		go
al		is		is
go	$\Rightarrow$	no	$\Rightarrow$	no
pe		of		of
to		pe		pa
co		pa		pe
to		th		th
th		ti		th
ai		to		th
of		to		ti
th		th		to
pa		th		to

2. How many letter comparisons would Boyer-Moore algorithm (with only the bad character rule) perform on the text a^n and pattern $a^{m-1}b$.

Answer: For the text a^n and pattern $a^{m-1}b$, Boyer-Moore algorithm (with only the bad character rule) would perform $n-(m-1) = n-m+1$ letter comparisons. This is because every time a pattern is aligned (starting at $index = 0$), the letter comparisons take place from right to left. As a result, the last character b would be compared against an a, and would result in a mismatch. However, since the rightmost a in the pattern is at position $m-1$, the number of alignments skipped = m-1-(m-1) = 0. Hence, the pattern is aligned at the next index position, and this is done for a total of $n-m+1$ alignments, resulting in a total of $n-m+1$ comparisons of b with a's in the text.

3. Give an example of a text and pattern that would result in the Boyer-Moore algorithm (with only the bad character rule) to perform N/M letter comparisons.
   Answer: The Boyer-Moore algorithm (with only the bad character rule) performs N/M letter comparisons for the `text = c c c c c c c c c` and the `pattern = a b d`.

4. Give an example of a text and pattern that would result in the worst case performance of the Rabin-Karp algorithm.
   Answer: The Rabin-Karp algorithm results in the worst case performance NM when the hash value computed for each x_i of the text, is the same and is also equal to hash computed for the pattern. The following example of `text = 1 1 1 1 1 1 1 1 1` and the `pattern = 1 1 1` results in the worst case performance for Rabin-Karp algorithm.

5. Perform substring search using the Rabin-Karp algorithm on the following the `text = 7 5 4 3 7 8 5 3 9 2 1 7`, and `pattern = 3 9 2`. The algorithm uses the following hash function: $h(x) = x \mod 23$.
   (a) When performing the substring search compute the hash functions for $x_4, \ldots x_{11}$ using rolling hash.
   (b) How many letter comparisons would the algorithm perform in this case?

   Answer: This is the details of running Rabin-Karp algorithm:

   Pattern Hash: 1
   Text Hash at 0: 18
   Text Hash = (18 + 23 - ((8 * 7) % 23))) % 23 = 8

Text Hash = ((8 * 10 + 3) % 23) = 14
Text Hash at 1: 14 Text Hash = (14 + 23 - ((8 * 5) % 23))) % 23 = 20
Text Hash = ((20 * 10 + 7) % 23) = 0
Text Hash at 2: 0 Text Hash = (0 + 23 - ((8 * 4) % 23))) % 23 = 14
Text Hash = ((14 * 10 + 8) % 23) = 10
Text Hash at 3: 10 Text Hash = (10 + 23 - ((8 * 3) % 23))) % 23 = 9
Text Hash = ((9 * 10 + 5) % 23) = 3
Text Hash at 4: 3 Text Hash = (3 + 23 - ((8 * 7) % 23))) % 23 = 16
Text Hash = ((16 * 10 + 3) % 23) = 2
Text Hash at 5: 2 Text Hash = (2 + 23 - ((8 * 8) % 23))) % 23 = 7
Text Hash = ((7 * 10 + 9) % 23) = 10
Text Hash at 6: 10 Text Hash = (10 + 23 - ((8 * 5) % 23))) % 23 = 16
Text Hash = ((16 * 10 + 2) % 23) = 1
Text Hash at 7: 1 Match!
Text Hash = (1 + 23 - ((8 * 3) % 23))) % 23 = 0
Text Hash = ((0 * 10 + 1) % 23) = 1
Text Hash at 8: 1 Text Hash = (1 + 23 - ((8 * 9) % 23))) % 23 = 21
Text Hash = ((21 * 10 + 7) % 23) = 10
Text Hash at 9: 10

text: 754378539217
pattern: 392

We would do 10 hash comparison in this case and then we 3 letter comparisons after we have the same hash for the pattern and the text at position 7.

6. What is the LZW encoding of the following input: T O B E O R N O T T O B E? You may assume that the input contains 7-bit ASCII characters and the output is in 8 bit codewords in hexadecimal. Provide the trie representing the symbol table. What is the compression ratio achieved?

Answer: The results of the encoding is as follow:

T	O	B	E	O	R	N	O	T	TO	BE	EoF
54	$4f$	42	45	$4f$	52	$4e$	$4f$	54	81	83	80

key	value
B	42
E	45
N	4e
O	4f
R	52
T	54
⋮	⋮
TO	81
OB	82
BE	83
EO	84
OR	85
RN	86
NO	87
OT	88
TT	89
TOB	8A

Trie of the symbol table is given in Figure 1.

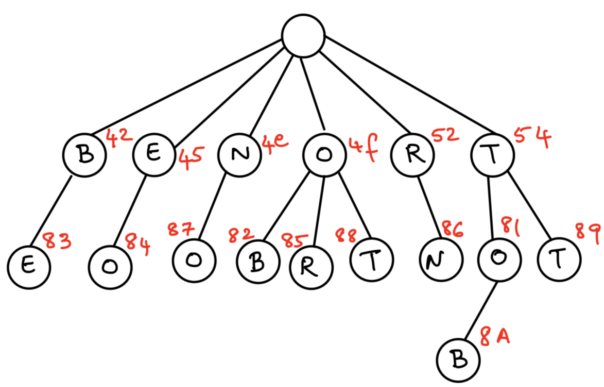

Figure 1: Trie symbol table for Question 6

Number of bits required to represent the encoding: $8 * 11 = 88$ (codeword for EoF not counted).

Number of bits required to represent the original data: $7 * 13 = 91$.
Compression rate: $\dfrac{88}{91} = 96\%$