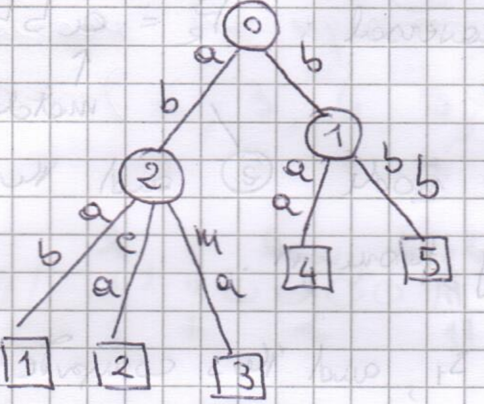
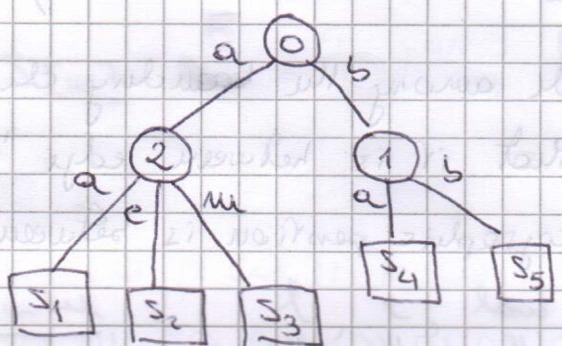
Solution 12/12/2023

Question #1

$S = \{abab, abca, abma, baa, bbb\}$



compacted Trie



Patricia Trie

Let us given the string $P_1 = aaa$, we search for its lexicographic position in S via three phases:

① downward Traversal: $P_1 = \underline{a}a\underline{a}$ and

match

reach the leaf s_1.

② Compute $lcp(s_1, P_1) = 1$

③ upward Traversal: we end up on the edge 

and since $P_1[2] = a < b = s_1[2]$ we have to go to the leftmost path and thus find that pattern $P_1 < s_1$.

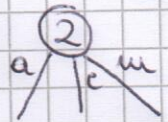
If, instead, we search for $P_2 = abb$, and repeat the 3 phases we get:

1. downward traversal: $P_2 = abb$ — match, mismatch

   so we stop at node 2 and thus take any leaf descending from it.

2. Say we take s_1, and thus compute $lcp(P_2, s_1) = 2$

3. upward traversal: stops at node 2 and



   thus we search among the branching chars for $P_2[3] = b$ and find that it is between edge "a" and edge "c" so the lexicographic position is between s_1 and s_2.

Question #2

$S = (2, 3, 4, 5, 6, 10, 11)$

- The $(2,6)$-dense code is therefore built on $2^b = 2 + 6 = 8$ $\Rightarrow b = 3$ bits. The first 2 configurations $\{000, 001\}$ are stoppers, the other 6 are continuers.

thus:

x	configuration (x)
0	000
1	001
2	010 000
3	010 001
4	011 000
5	011 001
6	100 000
7	100 001
8	101 000

x	config(x)
9	101 001
10	110 000
11	110 001

Let us now encode the sequence via Interpolative code.

$$S = (\underset{1}{2}, \underset{2}{3}, \underset{3}{4}, \underset{4}{5}, \underset{5}{6}, \underset{6}{10}, \underset{7}{11})$$

First number: $l = 1$, $r = 7$, $low = 2$, $high = 11$

$m = \frac{1+7}{2} = 4$

$S[m] = S[4] = 5$

range $= [low + m - l, high + m - r] = [5, 8]$

Encode $S[m] - \cancel{}\,5 = 0$ in $\lceil \lg_2 (8-5+1) \rceil = 2$ bits

$\Longrightarrow (00)_2$

Second number: $l = 1$, $r = 3$, $low = 2$, $high = 4$

by doing calculations you discover $\emptyset$ bits emitted

Third number: $l = 5$, $r = 7$, $low = 6$, $high = 11$

$m = \frac{5+7}{2} = 6$

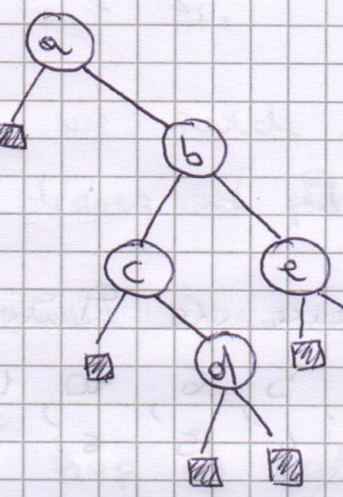
$S[m] = S[6] = 10$

range $= [6 + 6 - 5, 11 + 6 - 7] = [7, 10]$

Encode $S[6] - 7 = 3$ in $\lceil \lg_2 (10-7+1) \rceil = 2$ bits

$\Longrightarrow (11)_2$

Question #3



The binary tree consists of 5 nodes and 6 dummy leaves that are added for the succinct encoding.

$B = 1\,0\,1\,1\,1\,0\,1\,0\,0\,0\,0$ $\quad |B| = 11 = 2n+1$ where $n = 5$

We could think to store the labels in an additional array $L = [a, b, c, e, d]$ which is in bijection with the 1s of array B.

Question #4

$$L = 01\ 11\ 00\ 0101\ \overbrace{00}^{6^{th}}\ 11\ 11\ 00\ 11\ 00$$

$$H = 110\ 110\ 10\ 0\ \underset{\uparrow 6^{th}}{\overset{10}{10}}\ 10\ 10\ 110\ 00\ 10\ 00000$$

$n = 11 = \#1s$ in $H \Longrightarrow l = \frac{|L|}{n} = \frac{22}{11} = 2$ bits

$h = \lg_2(\#0s \text{ in } H) = \lg_2 16 = 4$ bits

Above I've indicated the 1 corresponding to the 6^{th} integer to be decoded, it is in the "group" corresponding to the configuration 4 (i.e., 0100) $= 10 - 6 = 4$

Combining (high, low)-configurations we get $(0100, 00) = 16$

Question # 5

T = BABABAC$ // I use capital letters for readibility.

sorted matrix

$ ~~BA~~BABAC
A BABAC$B
A BAC$BAB
A C$BABAB
B ABABAC$
B ABAC$BA
B AC$BABA
C $BABABA

$\Longrightarrow$ L = CBBB$AAA

position $ = 5 [preamble]

L' = CBBBAAA

MTF - step : [preamble]

$\mathcal{L} = \{A, B, C\} \longrightarrow 3$

$\mathcal{L} = \{C, A, B\} \longrightarrow 3\ 1\ 1$

$\mathcal{L} = \{B, C, A\} \longrightarrow 3\ 1\ 1$

3 3 1 1 3 1 1

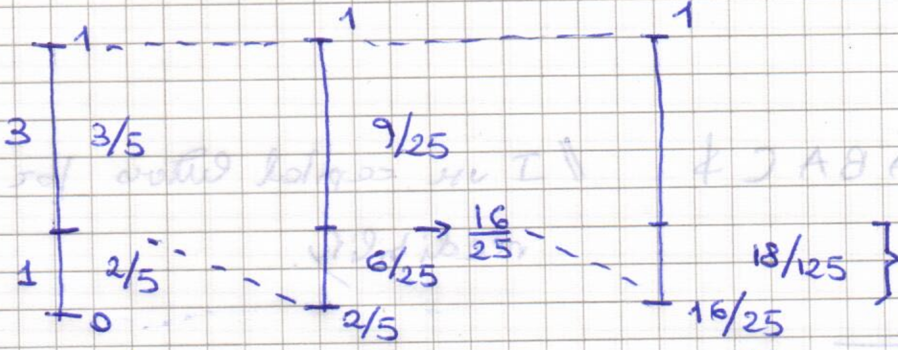
RLE Ø with Wheeler's code

Since the two runs of 1 have both length 2, they are encoded as $2+1 = 3 = (11)_2 \Rightarrow 1$.

therefore the output of this step is 3 3 1 3 1.

Arithmetic

$p(3) = \frac{3}{5}$ $p(1) = \frac{2}{5}$ [preamble]



This is the interval identified by the first three symbols.

$s_0 = 1$ $\quad s_1 = 3/5$ $\quad s_2 = 9/25$

$l_0 = 0$ $\quad l_1 = 2/5$ $\quad l_2 = \frac{16}{25}$

number to be encoded is : $\frac{16}{25} + \frac{1}{\cancel{2}} \cdot \frac{\cancel{18}^{9}}{125} = \frac{89}{125}$

$$\#\text{bits} = \left\lceil \lg_2 \frac{2}{s_3} \right\rceil = \left\lceil \lg_2 \frac{\cancel{2}}{\cancel{18}^{9}/125} \right\rceil = \left\lceil \lg_2 \frac{125}{9} \right\rceil = \left\lceil \lg_2 13.7 \right\rceil = 4$$

we encode $89/125$ in 4 bits, hence

$2 \cdot \frac{89}{125} = \frac{178}{125} \longrightarrow$ output 1 $\left(\frac{178}{125} - 1 = \frac{53}{125} \right)$

$2 \cdot \frac{53}{125} = \frac{106}{125} \longrightarrow$ output $\emptyset$

$2 \cdot \frac{106}{125} = \frac{212}{125} \longrightarrow$ output 1 $\left(\frac{212}{125} - 1 = 87 \right)$

$2 \cdot \frac{87}{125} = \frac{174}{125} \longrightarrow$ output 1

Question #6

Look at the book.