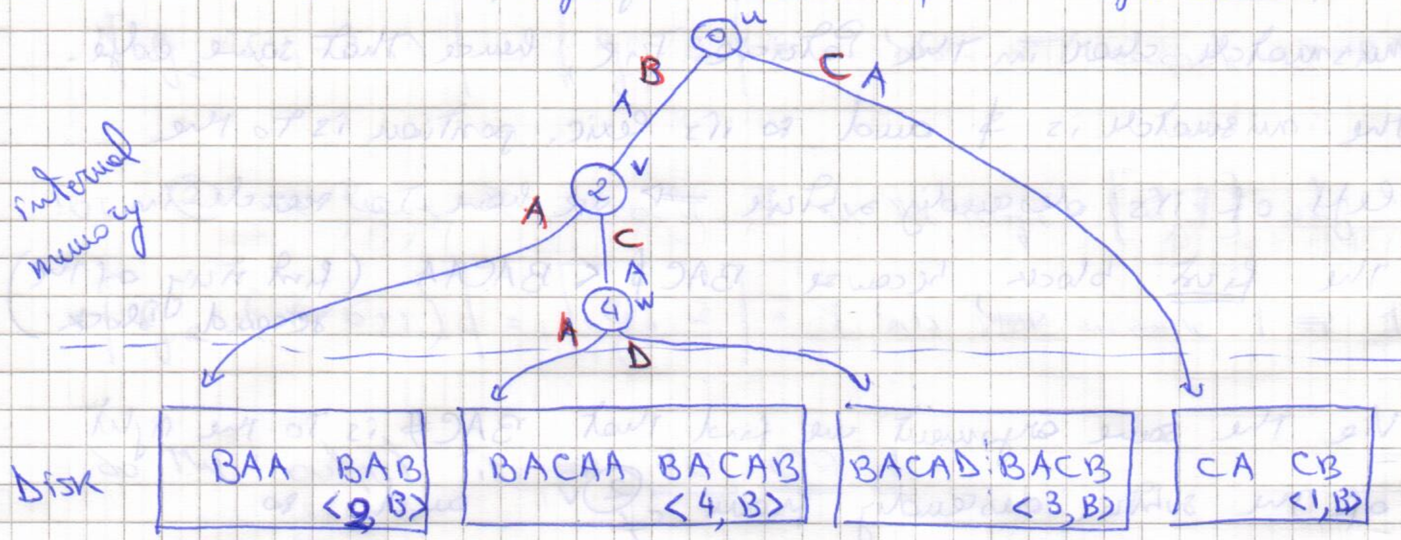
Algorithm Engineering
15 January 2024

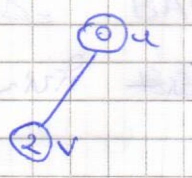
Q1

We divide the strings in blocks of 2 strings each



The red chars are the ones stored in the patricia trie.

- Let us now execute a lexicographic search for BB.
  This stops at node 2 v, picks any string descending, say BAA, computes LCP = 1 and then goes up stopping at edge



  . Since BAA < BB, the lexicographic position is to the right of that descending subtree.
  So we must search for it in the third block and find that it is at its end, so just before than the 4^{th} one.

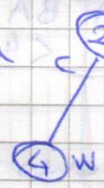
- Let us now execute the prefix search for BAC.
  We perform two searches (lexicographic) for BAC\$ and BAC# where \$ smaller, # larger } than any other char.

The lexicographic search for BAC\$ picks one of the strings descending from

, say BACAA.

Computes the LCP = 3 and then goes up to find the mismatch char in the Patricia Trie, hence that same edge. The mismatch is \$ and so its lexic. position is to the left of its descending subtrie, $\Rightarrow$ we have to search in the <u>first</u> block, because BAC\$ < BACAA (first string of the second block)

Via the same argument we find that BAC# is to the right of the subtrie descending from



and so we scan the <u>third</u> block.

In the first case, BAC\$ lies at the end of the first block. In the second case BAC# " " " " third block. Therefore the 2° and 3° blocks contain strings prefixed by BAC.

Q2 $l=1$, $low=2$, $r=3$, $high=7$ $\Rightarrow n=3$

pick element $m=2$, $S[2]=5$ and encode it in the range $[low+m-l,\ high+m-r] = [3, 6]$

using $\lceil \log_2 (6-3+1) \rceil = \log_2 4 = 2$ bits $\Rightarrow 5-3=2$

is encoded in 2 bits $\Rightarrow$ 10.

① the left call $(l = 1,\ low = 2,\ r = 1,\ high = 4)$

② the right call $(l = 3,\ low = 6,\ r = 3,\ high = 7)$

For ① we have to encode 2 in the range $[2, 4]$ using

$$\lceil \log_2 (4-2+1) \rceil = \lceil \log_2 3 \rceil = 2 \text{ bits} \Rightarrow \text{encode } 0 \Rightarrow 00$$

For ② we have to encode 7 in the range $[6, 7]$ using

$$\lceil \log_2 (7-6+1) \rceil = \lceil \log_2 2 \rceil = 1 \text{ bits} \Rightarrow \text{encode } 1 \Rightarrow 1.$$

So the output is $\Rightarrow$ $\underbrace{10}\ \underbrace{00}\ \underbrace{1}$

Q3

→BACAB
ABB
BBC
DD
DF

$S_<$ ABB $r = 1$

$S_=$ BACAB, BBC $r = 2$ ← $S_<$ BACAB, $S_>$ BBC

$S_>$ DD, DF $r = 1$ ← $S_=$ DD, DF $r = 2$ ← $S_<$ DD, $S_>$ DF

Q4

ABRABRA\$
·BRABRA\$A
RABRA\$AB
·ABRA\$ABR
·BRA\$ABRA
RA\$ - - - B
·A\$ - - - - - R
·\$ - - - - A

sort →

	L
\$	A
A\$	R
ABRA\$	R
ABRAA	\$ ← position 4
BRA\$	A
BRAB	A
RA\$	B
RAB	B

So the string passed to the next MTF step is ARRAABB

$\mathcal{L} = \{A, B, R\}$ $\quad$ ARRAABB $\longrightarrow$ 1, 3

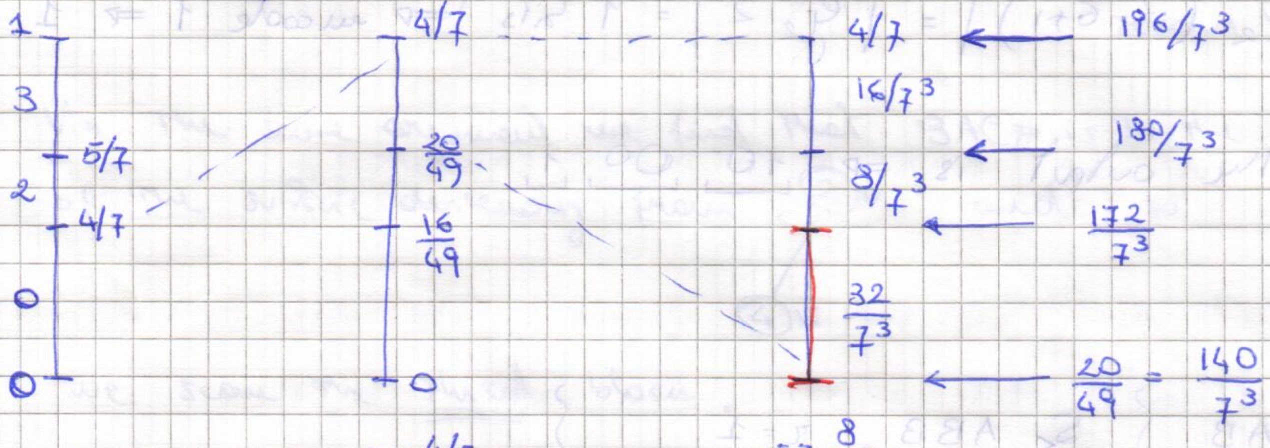
$\mathcal{L} = \{R, A, B\}$ $\quad$ RAABB $\longrightarrow$ 1, 2

$\mathcal{L} = \{A, R, B\}$ $\quad$ ABB $\longrightarrow$ 1, 3, 1

We have therefore to encode with RLE0: 1, 3, 1, 2, 1, 3, 1

which becomes, with the Wheeler code: 0, 3, 0, 2, 0, 3, 0

$P(0) = \frac{4}{7}$, $P(1) = 0$, $P(2) = \frac{1}{7}$, $P(3) = \frac{2}{7}$



$s = 1$ $\quad$ $s = 4/7$ $\quad$ $s = \frac{8}{49}$

The number to be encoded is: $\frac{140}{7^3} + \frac{16}{7^3} = \frac{156}{7^3} = \frac{156}{343}$

The number of bits to be emitted is $= \left\lceil \lg_2 \frac{2}{s} \right\rceil = \left\lceil \lg_2 \frac{2}{\frac{32}{7^3}} \right\rceil =$

$= \left\lceil \lg_2 \frac{7^3}{16} \right\rceil = \left\lceil \lg_2 21,\ldots \right\rceil = 5$ bits $\Rightarrow$ then you apply

the procedure Convert to $\frac{156}{343}$ and emit 5 bits.

Q5

	h	h1	h2
AA	0	3	3
AB	1	4	6
BB	2	6	-
BC	3	7	-
CC	4	9	-

$\longleftarrow$ this creates a loop and thus the algorithm stops.

Anyway, the system of equations has a solution, even for $g(3) = 0$.