

Information Retrieval

5/2/21

Q1. $S = (1, 1, 3, 2, 2, 4, 2, 1, 3)$

$$h_1(x) = 2x \pmod 7$$

$$h_2(x) = 3x \pmod 7$$

BF =

0	1	2	3	4	5	6
0	1	5	3	3	1	5

< key, value cell >

- <4,1>
- <1,1> <1,1>
- <2,1> <4,1> <3,1>
- <1,2> <1,2> <2,2>
- <3,3> <1,3> <2,3>
- <2,2>
- <2,3>
- <2,4>
- <3,5>
- <1,4>
- <3,5>

Query (2) = $\min(BF[4], BF[6]) = \min(3, 5) = 3$

The value is correct because key 2 has no conflicts on cell BF[4].

Q2.

- 10 → 3, 10, 11, 13, 15, 17
 - 11 → 5, 10, 11, 13, 14
 - 12 → 3, 10, 11, 13, 21, 25, 30
- ← copies from
- copies from 10 because it copies 4 items

node	Out of	ref	copy list	Extra nodes
11	5	1	0011100	5, 14
12	7	2	1111100	21, 25, 30

node	outd	ref.	#blocks	Copy blocks	Extra
11	5	1	3	0, 0, 2, x	5, 14
12	7	2	2	1, 3, 1	21, 25, 30

↑
front bit

Q3

• $A = 01000$

$A' = 01001 \rightarrow FC(A') = \langle 4, 1 \rangle$ or $\langle 5, 4, 1 \rangle$ ← but original string

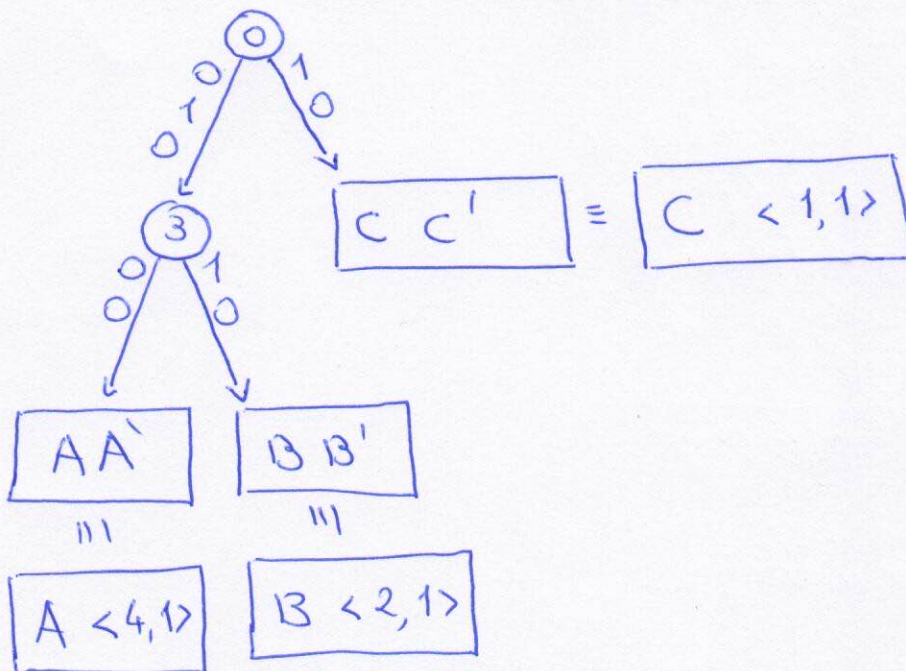
• $B = 01010$

$B' = 01101 \rightarrow FC(B') = \langle 2, 1 \rangle$ or $\langle 5, 2, 1 \rangle$

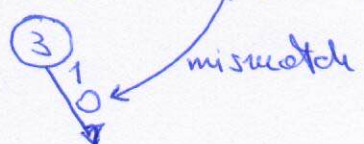
• $C = 10$

$C' = 11 \rightarrow FC(C') = \langle 1, 1 \rangle$ or $\langle 2, 1, 1 \rangle$

Build the Compressed Trie on A, B, C



Search for $S = 01011$, we visit the compressed trie up to the edge



we find the mismatch which shows that $B < S$ so the search for S must continue in B 's block.

Then the search algorithm jumps in the disk page BB'

and scans it searching for S. Then it discovers that S does not belong to the set of strings.

Q4.

$$F_{\text{-new}} = \begin{array}{c} x _ a b a b \\ \hline \downarrow h \quad \downarrow \quad \downarrow \\ 0 \quad 3 \quad 3 \end{array}$$

$$h(c_1 c_2) = (c_1 + c_2) \bmod 7$$

$$a = 1, b = 2, _ = 3, x = 4$$

$$F_{\text{-old}} = \begin{array}{c} \quad \quad 6 \quad \quad 5 \\ \quad \quad \downarrow \quad \downarrow \\ a \ b \ x \ x \ a \ b \\ \hline \downarrow \quad \quad \downarrow \quad \downarrow \\ 3 \quad \quad 1 \quad 3 \end{array}$$

Client $\xrightarrow{011}$ Server, indicating that the second and third blocks of $F_{\text{-new}}$ have been identified in $F_{\text{-old}}$ and they are equal to "ab".

The server answers:

$$\text{flip}(abab \mid x _) = \langle 0, 0, x \rangle \langle 0, 0, _ \rangle$$

Q5.

And. not (t_1, t_2)

$x = \text{head}(t_1);$

$y = \text{head}(t_2);$

while $(x \neq \text{NULL})$ do

if $(y = \text{NULL}) \parallel (\text{doc}[x] < \text{doc}[y])$

{ print doc[x]; $x = \text{next}[x];$ }

else $y = \text{next}[y]$