Q.1.(a) Solution -

The given inverted index is -

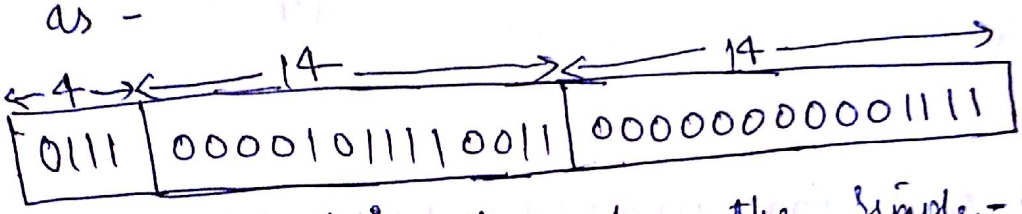
(755, 15), (1065, 31), (1280, 32), (2100, 12)

Now, let's define the d-gap representation for above indices -

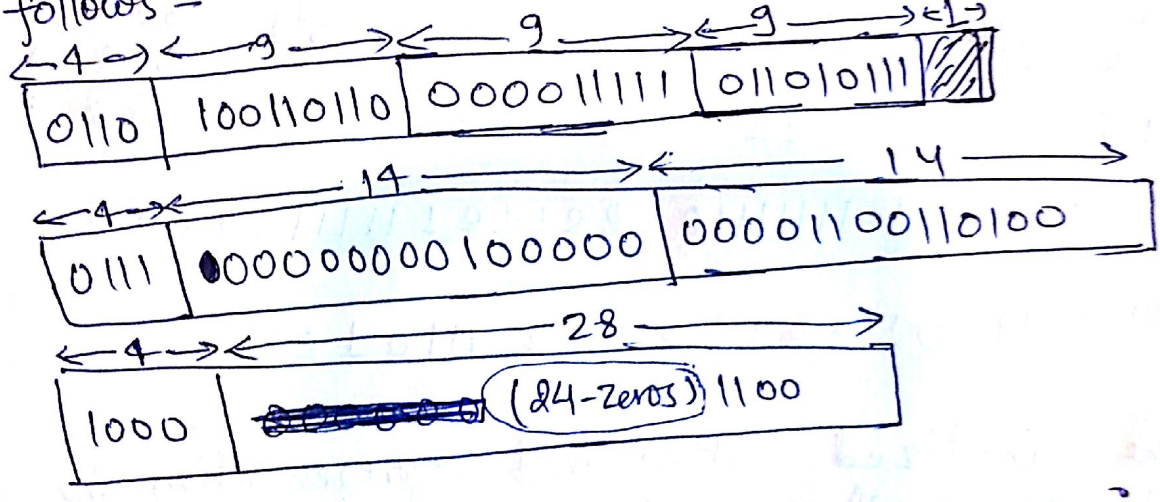
(755, 15), (310, 31), (215, 32), (820, 12)

10-bits 4-bits 9-bits 5-bits 8-bits 6-bits 10-bits 4-bits

Now, we use code length of 14 ~~bits~~ for two numbers as -



Then we continuously use the Simple-9 method as follows -



from this, we need 16 bytes by using Simple-9 compression method, whereas ~~this~~ without compression we may need 24 bytes.

Q.1.(b) Solution

The given inverted-index is –

(755, 15), (1065, 31), (1280, 32), (2100, 12)

The d-gap representation can be written as –

755 15 310 31 215 32 820 12

Now, we use bit-aligned method to compress the above representation as –

$755 - 512 \rightarrow 2^{9}$, remainder 243

$15 - 8 \rightarrow 2^{3}$, remainder 7 (111)

$310 - 256 \rightarrow 2^{8}$, remainder 44

$\Rightarrow$ 1111111110 0 11 11 00 11 1110 1 1 1 1111 11110 0 0 1 0 1 1 0 0

$31 - 16 - 2^{4}$, remainder 15

$215 - 128 \rightarrow 2^{7}$, remainder 77

$32 - 32 \rightarrow 2^{5}$, remainder 0

$820 - 512 \rightarrow 2^{9}$, remainder 308

$12 - 8 \rightarrow 2^{3}$, remainder 4

$\Rightarrow$ 1111 0 1 1 1 11111110 1 0 0 1 1 0 1 111110 0 0 0 0 0

$\Rightarrow$ 111111111 0 1 0 0 1 1 0 1 0 0 1110 1 0 0

So, the finalized bit-aligned representation for above inverted index will be –

1 1111 1111 0 0 1111 0011 1110 1111 1111 1110 0010 1100 1111 0111 1110 0000 01 1111 1111 0100 1101 0011 10100

So, using the bit-aligned method we need 104 bits $\cong$ 13 bytes.