CMSC 476676 Optional Homework V-Byte Encoding

What do the following V-byte Encodings represent?

10010000000010011000010010000101

10100011000100011001001110010100

11000011001000011000001010000110

Answers

1.

a.16

b.1156

c.5

2.

a.17

b.2195

c.20

3.

a.33

b.4098

c.6

Compute big O (order of complexity) for the work required to compute the pairwise similarity scores for 10^5 docs, 10^6 terms, ~10^4 terms/doc using:

4. Brute Force

a. O(\_\_10^14\_\_\_\_\_)

5. Doc-at-a-time

a. O(\_\_\_10^13\_\_\_\_)

6. Term-at-a-time

a. O(\_\_\_10^12\_\_\_\_)