# Problem 5 : (24 points);

Suppose you have a collection of 5 documents, and only 10 terms are used:

Term1 Term2 Term3 Term4 Term5 Term6 Term7 Term8 Term9 Term10

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

DOC1 0 3 2 4 0 5 0 0 4 2

DOC2 3 0 1 4 3 0 0 5 1 6

DOC3 6 0 5 1 2 0 2 5 0 7

DOC4 1 8 0 2 0 1 6 0 2 1

DOC5 2 7 0 0 0 3 0 2 3 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

List the values of the gaps for the last three terms in your index computed for this collection. (**1 point**)

Encode these gaps with (i) unary codes (**3 points**); (ii) Gamma codes (**10 points**); and (iii) Delta codes (**10 points**). You are allowed to write a program to enable you computing the codes. Please add to the exam the code of the program if you chose to use one.

SOLUTION 2.I:

Index Computation:

*term-->DOCid*

term1--> [2, 3, 4, 5]

term2--> [1, 4, 5]

term3--> [1, 2, 3]

term4--> [1, 2, 3, 4]

term5--> [2, 3]

term6--> [1, 4, 5]

term7--> [3, 4]

term8--> [2, 3, 5]

term9--> [1, 2, 4, 5]

term10--> [1, 2, 3, 4]

Gaps for Term 8 and the unary codes for the gaps:

Gap is calculated as difference of current value from its previous value

Gaps for Term 8 : [2, 1, 2]

Unary code for n is n1’s followed by 0

Unary code for 2 : 110

Unary code for 1 : 10

Unary code for 2 : 110

Unary code of Gaps for Term 8: [110,10,110]

Gaps for Term 9 and the unary codes for the gaps:

Gap is calculated as difference of current value from its previous value

Gaps for Term 9 : [1, 1, 2, 1]

Unary code for n is n1’s followed by 0

Unary code for 1 : 10

Unary code for 2 : 110

Unary code for 1 : 10

Unary code for 1 : 10

Unary code of Gaps for Term 9: [10,110,10,10]

Gaps for Term 10 and the unary codes for the gaps:

Gap is calculated as difference of current value from its previous value

Gaps for Term 10 : [1, 1, 1, 1]

Unary code for n is n1’s followed by 0

Unary code for 1 : 10

Unary code for 1 : 10

Unary code for 1 : 10

Unary code for 1 : 10

Unary code of Gaps for Term 10: [10,10,10,10]

SOLUTION 2.II:

Gamma codes for the gaps in the posting files of Term 8:

Gamma code of Gap 2 :

a. Binary representation of Gap = 10

Offset = remove the leading 1 from binary representation of Gap

= 0

b. Length of offset = 1

Encoding length in unary code = 10

Gamma code of Gap 2 = Concatenation of length and offset

= 100

Gamma code of Gap 1 :

a. Binary representation of Gap = 1

Offset = remove the leading 1 from binary representation of Gap

= “ ”

b. Length of offset = 0

Encoding length in unary code = 0

Gamma code of Gap 1 = Concatenation of length and offset

= 0

Gamma code of Gap 2 :

a. Binary representation of Gap = 10

Offset = remove the leading 1 from binary representation of Gap

= 0

b. Length of offset = 1

Encoding length in unary code = 10

Gamma code of Gap 2 = Concatenation of length and offset

= 100

Gamma code for Gaps of term 8:

100

0

100

Gamma codes for the gaps in the posting files of Term 9:

Gamma code of Gap 1 :

a. Binary representation of Gap = 1

Offset = remove the leading 1 from binary representation of Gap

= “ ”

b. Length of offset = 0

Encoding length in unary code = 0

Gamma code of Gap 1 = Concatenation of length and offset

= 0

Gamma code of Gap 1 :

a. Binary representation of Gap = 1

Offset = remove the leading 1 from binary representation of Gap

= “ ”

b. Length of offset = 0

Encoding length in unary code = 0

Gamma code of Gap 1 = Concatenation of length and offset

= 0

Gamma code of Gap 2 :

a. Binary representation of Gap = 10

Offset = remove the leading 1 from binary representation of Gap

= 0

b. Length of offset = 1

Encoding length in unary code = 10

Gamma code of Gap 2 = Concatenation of length and offset

= 100

Gamma code of Gap 1 :

a. Binary representation of Gap = 1

Offset = remove the leading 1 from binary representation of Gap

= “ ”

b. Length of offset = 0

Encoding length in unary code = 0

Gamma code of Gap 1 = Concatenation of length and offset

= 0

Gamma code for Gaps of term 9:

0

0

100

0

Gamma codes for the gaps in the posting files of Term 10:

Gamma code of Gap 1 :

a. Binary representation of Gap = 1

Offset = remove the leading 1 from binary representation of Gap

= “ ”

b. Length of offset = 0

Encoding length in unary code = 0

Gamma code of Gap 1 = Concatenation of length and offset

= 0

Gamma code of Gap 1 :

a. Binary representation of Gap = 1

Offset = remove the leading 1 from binary representation of Gap

= “ ”

b. Length of offset = 0

Encoding length in unary code = 0

Gamma code of Gap 1 = Concatenation of length and offset

= 0

Gamma code of Gap 1 :

a. Binary representation of Gap = 1

Offset = remove the leading 1 from binary representation of Gap

= “ ”

b. Length of offset = 0

Encoding length in unary code = 0

Gamma code of Gap 1 = Concatenation of length and offset

= 0

Gamma code of Gap 1 :

a. Binary representation of Gap = 1

Offset = remove the leading 1 from binary representation of Gap

= “ ”

b. Length of offset = 0

Encoding length in unary code = 0

Gamma code of Gap 1 = Concatenation of length and offset

= 0

Gamma code for Gaps of term 10:

0

0

0

0

SOLUTION 2.III:

Delta codes for the gaps in the posting files of Term 8:

🡪Delta code for Gap 2:

a. Binary representation of Gap = 10

Offset = Remove leading 1 from binary representation of Gap

= 0

b. Length of Binary Representation of Gap = 2

Encode length in Gamma code :

Gamma code of 2 :

a. Binary representation of 2 = 10

Offset = remove the leading 1 from binary representation of 2

= 0

b. Length of offset = 1

Encoding length in unary code = 10

Gamma code of 2 = Concatenation of length and offset

= 100

Delta code of Gap 2 = Concatenation of length and offset

= 1000

🡪Delta code for Gap 1:

a. Binary representation of Gap = 1

Offset = Remove leading 1 from binary representation of Gap

= “ “

b. Length of Binary Representation of Gap = 1

Encode length in Gamma code :

Gamma code of 1 :

a. Binary representation of 1 = 1

Offset = remove the leading 1 from binary representation of 1

= “ ”

b. Length of offset = 0

Encoding length in unary code = 0

Gamma code of 1 = Concatenation of length and offset

= 0

Delta code of Gap 1 = Concatenation of length and offset

= 0

🡪Delta code for Gap 2:

a. Binary representation of Gap = 10

Offset = Remove leading 1 from binary representation of Gap

= 0

b. Length of Binary Representation of Gap = 2

Encode length in Gamma code :

Gamma code of 2 :

a. Binary representation of 2 = 10

Offset = remove the leading 1 from binary representation of 2

= 0

b. Length of offset = 1

Encoding length in unary code = 10

Gamma code of 2 = Concatenation of length and offset

= 100

Delta code of Gap 2 = Concatenation of length and offset

= 1000

Delta codes for Gaps in posting files of term 8:

1000

0

1000

Delta codes for the gaps in the posting files of Term 9:

🡪Delta code for Gap 1:

a. Binary representation of Gap = 1

Offset = Remove leading 1 from binary representation of Gap

= “ “

b. Length of Binary Representation of Gap = 1

Encode length in Gamma code :

Gamma code of 1 :

a. Binary representation of 1 = 1

Offset = remove the leading 1 from binary representation of 1

= “ ”

b. Length of offset = 0

Encoding length in unary code = 0

Gamma code of 1 = Concatenation of length and offset

= 0

Delta code of Gap 1 = Concatenation of length and offset

= 0

🡪Delta code for Gap 1:

a. Binary representation of Gap = 1

Offset = Remove leading 1 from binary representation of Gap

= “ “

b. Length of Binary Representation of Gap = 1

Encode length in Gamma code :

Gamma code of 1 :

a. Binary representation of 1 = 1

Offset = remove the leading 1 from binary representation of 1

= “ ”

b. Length of offset = 0

Encoding length in unary code = 0

Gamma code of 1 = Concatenation of length and offset

= 0

Delta code of Gap 1 = Concatenation of length and offset

= 0

🡪Delta code for Gap 2:

a. Binary representation of Gap = 10

Offset = Remove leading 1 from binary representation of Gap

= 0

b. Length of Binary Representation of Gap = 2

Encode length in Gamma code :

Gamma code of 2 :

a. Binary representation of 2 = 10

Offset = remove the leading 1 from binary representation of 2

= 0

b. Length of offset = 1

Encoding length in unary code = 10

Gamma code of 2 = Concatenation of length and offset

= 100

Delta code of Gap 2 = Concatenation of length and offset

= 1000

🡪Delta code for Gap 1:

a. Binary representation of Gap = 1

Offset = Remove leading 1 from binary representation of Gap

= “ “

b. Length of Binary Representation of Gap = 1

Encode length in Gamma code :

Gamma code of 1 :

a. Binary representation of 1 = 1

Offset = remove the leading 1 from binary representation of 1

= “ ”

b. Length of offset = 0

Encoding length in unary code = 0

Gamma code of 1 = Concatenation of length and offset

= 0

Delta code of Gap 1 = Concatenation of length and offset

= 0

Delta codes for Gaps in posting files of term 9:

0

0

1000

0

Delta codes for the gaps in the posting files of Term 10:

🡪Delta code for Gap 1:

a. Binary representation of Gap = 1

Offset = Remove leading 1 from binary representation of Gap

= “ “

b. Length of Binary Representation of Gap = 1

Encode length in Gamma code :

Gamma code of 1 :

a. Binary representation of 1 = 1

Offset = remove the leading 1 from binary representation of 1

= “ ”

b. Length of offset = 0

Encoding length in unary code = 0

Gamma code of 1 = Concatenation of length and offset

= 0

Delta code of Gap 1 = Concatenation of length and offset

= 0

🡪Delta code for Gap 1:

a. Binary representation of Gap = 1

Offset = Remove leading 1 from binary representation of Gap

= “ “

b. Length of Binary Representation of Gap = 1

Encode length in Gamma code :

Gamma code of 1 :

a. Binary representation of 1 = 1

Offset = remove the leading 1 from binary representation of 1

= “ ”

b. Length of offset = 0

Encoding length in unary code = 0

Gamma code of 1 = Concatenation of length and offset

= 0

Delta code of Gap 1 = Concatenation of length and offset

= 0

🡪Delta code for Gap 1:

a. Binary representation of Gap = 1

Offset = Remove leading 1 from binary representation of Gap

= “ “

b. Length of Binary Representation of Gap = 1

Encode length in Gamma code :

Gamma code of 1 :

a. Binary representation of 1 = 1

Offset = remove the leading 1 from binary representation of 1

= “ ”

b. Length of offset = 0

Encoding length in unary code = 0

Gamma code of 1 = Concatenation of length and offset

= 0

Delta code of Gap 1 = Concatenation of length and offset

= 0

🡪Delta code for Gap 1:

a. Binary representation of Gap = 1

Offset = Remove leading 1 from binary representation of Gap

= “ “

b. Length of Binary Representation of Gap = 1

Encode length in Gamma code :

Gamma code of 1 :

a. Binary representation of 1 = 1

Offset = remove the leading 1 from binary representation of 1

= “ ”

b. Length of offset = 0

Encoding length in unary code = 0

Gamma code of 1 = Concatenation of length and offset

= 0

Delta code of Gap 1 = Concatenation of length and offset

= 0

Delta codes for Gaps in posting files of term 9:

0

0

0

0