Student Name: _____SAMPLE SOLUTION_____   Student Number: _____

Circle check and/or write the correct answer (12 in total):

1- According to Heaps law the number of terms continues to increase with the increase of the corpus (collection of documents) size, but the rate of change may vary from one language to another.
□**A True**, □False

2- According to Zipf's law if the frequency of the 3 ranked term is 10,000 then the frequency of the sixth ranked term is: □**A: 20,000** □**B:5,000,** □**C: 3,333,** □ **D: None of the above** ➔ **Answer is: ___B___**

3- When using a full bytes in variable byte coding of GAPs the amount of waste is:
□**A: 50%** □**B:12.5%,** □**C: 25%,** □ **D: None of the above** ➔   **Answer is: ___B_____**

4- The following statement is most likely to be   □A True, □**False**
"**The average TOKEN length in a corpus is usually larger than the average TERM length**"

5- Every document has more terms than tokens: □A True, □**False**

6- To code the document number in a collection of 400,000 documents we need:
□**A: 20 bits** □**B:19 bits,** □**C: 21 bits,** □ **D: 18 bits,** □ **E:None of the above** ➔ **Answer is: ____ B ___**

7- The following is a valid gamma code: **110011:**   □A True, □**False**

8- The following is a valid gamma code: **1110111:**   □**A True**, □False

9- The gamma code for the sequence of gaps 4, 7, 13 is: Answer is: 11000110111110101
10- The following gamma code represents the following sequence of decimal gaps:

   1110000110011110111   Answer is: ___8,5,15_____