Close

Week 2 Quiz

10 questions

1  
point

Correct

1 / 1 points

1.

Let w1, w2, and w3 represent three words in the dictionary of an inverted index. Suppose we have the following document frequency distribution:

|  |  |
| --- | --- |
| **Word** | **Document Frequency** |
| w1 | 1000 |
| w2 | 100 |
| w3 | 10 |

Assume that each posting entry of document ID and term frequency takes exactly the same disk space. Which word, if removed from the inverted index, will save the **most** disk space?



w1



We cannot tell from the given information.



w3



w2

Incorrect

0 / 1 points

2.

Assume we have the same scenario as in Question 1. If we enter the query Q= “w1 w2” then the **minimum** possible number of accumulators needed to score all the matching documents is:



1000



100



10



1100

Correct

1 / 1 points

3.

The gamma code for the term frequency of a certain document is **1110010**. What is the term



10



11



12



9

Correct

1 / 1 points

4.

When using an inverted index for scoring documents for queries, a shorter query always uses fewer score accumulators than a longer query.



False



True

Correct

1 / 1 points

5.

What is the advantage of tokenization (normalize and stemming) before index?



Improves performance by mapping words with similar meanings into the same indexing term



Extracts words as lexical units from strings of text



Reduces the number of terms (size of vocabulary)

Correct

1 / 1 points

6.

What can't an inverted index alone do for fast search?



Retrieve documents that are relevant to the query



Search document contains "A" or "B"



Search document contains "A" and "B"

Correct

1 / 1 points

7.

If Zipf's law does not hold, will an inverted index be much faster or slower?



Faster



Slower

Correct

1 / 1 points

8.

In BM25, the TF after transformation has upper bound



1



k +1



k

Incorrect

0 / 1 points

9.

Which of the following are weighing heuristics for the vector space model?



IDF weighting



Document length normalization



TF weighting and transformation

Correct

1 / 1 points

10.

Which of the following integer compression has equal-length coding?



Unary



Binary



*γ*-code

Submit Quiz