Close

Week 2 Practice Quiz

**10/10** points earned (100%)

Excellent!

Retake

[Next](https://www.coursera.org/learn/text-retrieval/exam/kHw7H/week-2-quiz)

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 | 1 |
| w2 | 5 |
| w3 | 10 |

Assume that each posting entry of document ID and term frequency takes exactly the same disk space. Which word's postings list will occupy the largest disk space?



w3

**Correct Response**

Explanation: The postings list of w3 has the largest number of entries and thus occupies the largest space.



w2



w1

Correct

1 / 1 points

2.

Assume we have the same scenario as in Question 1. If we enter a query Q= “w1 w2 w3” then the **maximum** possible number of accumulators needed to score all the matching documents is:



10



1



16

**Correct Response**

Explanation: If the three postings lists are mutually exclusive (have no common elements), then we will have 16 unique documents each matching exactly 1 of the query terms.



5

Correct

1 / 1 points

3.

Assume that the d-gap between two documents is equal to 9. If you want to compress this d-gap with a **gamma** code, what will be the binary representation of the code?



1110001

**Correct Response**

Explanation: 1+floor(log(9)) = 4, which can be represented as 1110 in unary code. 9 – 2^(floor(log(9))) = 1, which can be represented as 001 in a uniform code with 3 bits. The gamma code is the concatenation of the unary and uniform codes.



1110010



1110000



1110011

Correct

1 / 1 points

4.

Why is TF transformation needed?



So that computation is more efficient



To capture the intuition of "diminishing return" from higher TF

**Correct Response**

Correct

1 / 1 points

5.

What is the upperbound for BM25 transformation?



k+1

**Correct Response**



k-1



k

Correct

1 / 1 points

6.

Do we always want to penalize a long document?



Yes



No

**Correct Response**

If it uses more words, then we want to penalize more, but if it has more content, then we want to penalize less.

Correct

1 / 1 points

7.

Which is true about pivoted length normalization?



It always rewards.



It always penalizes.



It has both a penalization and reward effect.

**Correct Response**

Correct

1 / 1 points

8.

Is word segmentation on Chinese easier than English?



Yes



No

**Correct Response**

Correct

1 / 1 points

9.

What is NOT the advantage for using inverted index?



Inverted index can map words of the same meaning into one slot.

**Correct Response**



It is more efficient than sequentially scanning docs.



It can search for documents that contains both "A" and "B" efficiently.

Correct

1 / 1 points

10.

What does Zipf's law tell you?



Words are evenly distributed.



There are only a few words that have a small probability.



There are many words that have a small probability.

**Correct Response**