Congratulations! You passed!

TO PASS 70% or higher

Keep Learning

GRADE
90%

Week 2 Practice Quiz

TOTAL POINTS 10

1. Let w_1, w_2, and w_3 represent three words in the dictionary of an inverted index. Suppose we have the following document frequency distribution: **1 / 1 point**

Word	Document Frequency
w_1	1
w_2	5
w_3	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?

- w_1
- (selected) w_3
- w_2

Correct

Explanation: The postings list of w_3 has the largest number of entries and thus occupies the largest space.

2. Assume we have the same scenario as in Question 1. If we enter a query Q= "w_1 w_2 w_3" then the **maximum** possible number of accumulators needed to score all the matching documents is: **1 / 1 point**

- 10
- 5

- (●) 16
- ( ) 1

✓ **Correct**

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.

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? **1 / 1 point**

- (●) 1110001
- ( ) 1110000
- ( ) 1110011
- ( ) 1110010

✓ **Correct**

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.

4. Why is TF transformation needed? **1 / 1 point**

- ( ) So that computation is more efficient
- (●) To capture the intuition of "diminishing return" from higher TF

✓ **Correct**

5. What is the upperbound for BM25 transformation? **1 / 1 point**

- (●) k+1
- ( ) k
- ( ) k-1

✓ **Correct**

6. Do we always want to penalize a long document? **1 / 1 point**

- (●) No
- ( ) Yes

✓ **Correct**

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

7. Which is true about pivoted length normalization? **0 / 1 point**

- (●) It always penalizes.
- ( ) It has both a penalization and reward effect.
- ( ) It always rewards.

! **Incorrect**

8. Is word segmentation on Chinese easier than English? **1 / 1 point**

- (●) No
- ( ) Yes

Correct

Correct

9. What is NOT the advantage for using inverted index? **1 / 1 point**

- (●) Inverted index can map words of the same meaning into one slot.
- ( ) It is more efficient than sequentially scanning docs.
- ( ) It can search for documents that contains both "A" and "B" efficiently.

✓ **Correct**

10. What does Zipf's law tell you? **1 / 1 point**

- (●) There are many words that have a small probability.
- ( ) There are only a few words that have a small probability.
- ( ) Words are evenly distributed.

✓ **Correct**