Feedback — Week 2 Practice Quiz

Help Center

Thank you. Your submission for this quiz was received.

You submitted this quiz on **Mon 30 Mar 2015 11:04 AM PDT**. You got a score of **7.00** out of **7.00**.

Question 1

Suppose a query has a total of 5 relevant documents in a collection of 100 documents. System A and System B have each retrieved 10 documents, and the relevance status of the ranked lists is shown below:

```
System A: [+ + - - - - - - - -]
System B: [- + - - + - - - - +]
```

where the leftmost entry corresponds to the highest ranked document, and the rightmost entry corresponds to the lowest ranked document. A "+" indicates a relevant document and a "-" corresponds to a non-relevant one. For example, the top ranked document retrieved by System A is relevant whereas the top ranked one by B is non-relevant.

What is the **precision at 10 documents** of both systems?

Your Answer		Score	Explanation
◉ P(A) = 2/10 P(B)= 3/10	✔	1.00	
○ P(A) = 2/100 P(B)= 3/100			
○ P(A) = 2/5 P(B)= 3/5			
○ P(A) = 8/100 P(B)= 7/100			
Total		1.00 / 1.00	

Question 2

Assume the same scenario as in Question 1. What is the **recall** of both systems?

Your Answer		Score	Explanation
R(A) = 8/100 R(B)= 7/100			
R(A) = 2/10 R(B)= 3/10			
R(A) = 2/5 R(B)= 3/5	✔	1.00	
R(A) = 2/100 R(B)= 3/100			
Total		1.00 / 1.00	

Question 3

Assume the same scenario as in Question 1. What is the **average precision** of both systems?

Your Answer		Score	Explanation
AP(A) = 2/10 AP(B) = 3/25			
AP(A) = 3/10 AP(B) = 9/20			
AP(A) = 2/100 AP(B) = 3/250			
AP(A) = 2/5 AP(B) = 6/25	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

Explanation: The average precision is the sum of the precisions at each time a relevant document is retrieved **divided by the total number of relevant documents**. AP(A) = (1/1 + 2/2)/5 = 2/5 and AP(B) = (1/2 + 2/5 + 3/10)/5 = 6/25

Question 4

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:

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?

Your Answer		Score	Explanation
◉ w_3	✔	1.00	
○ w_1			
○ w_2			
Total		1.00 / 1.00	

Question Explanation

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

Question 5

Assume we have the same scenario as in Question 4. 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:

Your Answer		Score	Explanation
○ 5			
◉ 16	✔	1.00	
○ 10			
○ 1			
Total		1.00 / 1.00	

Question Explanation

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

Question 6

Assume that 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?

Your Answer		Score	Explanation
1110000			
1110010			
1110011			
1110001	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

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.

Question 7

Assume you have two retrieval systems X and Y. If X has a higher MAP (mean average precision), can Y have a higher gMAP (geometric mean average precision)?

Your Answer		Score	Explanation
Yes	✔	1.00	
No			
Total		1.00 / 1.00	

Question Explanation

Explanation: This is possible. For example, if both systems are being evaluated on two queries where the average precisions for system X are $AP_x(Q_1) = 0.9$ and $AP_X(Q_2) = 0.01$, and those for system Y are $AP_Y(Q_1) = 0.2$ and $AP_Y(Q_2) = 0.2$. Clearly, X has a higher MAP; MAP_x =

(0.9+0.01)/2 = 0.455 and MAP_Y = (0.2+0.2)/2 = 0.2

Recall that the geometric mean of two values is the square root of their product. So $gMAP_x$ = sqrt(0.9*0.01) = 0.094 and $gMAP_Y$ = sqrt(0.2*0.2) = 0.2. This illustrates an important property of gMAP: It is dominated by the low values of average precision, making it a good indicator of how the system performs in the presence of “hard” queries.