

START OF QUIZ

Student ID:

88683578,Mao,MingJia

Question 1

Topic: Lecture 7

Source: Lecture 7

Why do we generally care more about precision than recall in IR? (1)

Question 2

Topic: Lecture 5

Source: Lecture 5

Why do we need methods like t-SNE? (1)

Question 3

Topic: Lecture 6

Source: Lecture 6

Why do we need a "human in the loop" for topic modeling? (1)

Question 4

Topic: Lecture 7

Source: Lecture 7

Explain why the cosine similarity between a document and query vector is roughly equivalent to adding up the TF-IDF scores of each word in the document that occurs in the query.
(2)

Question 5

Topic: Lecture 8

Source: Lecture 8

In class, I mentioned that high k value for BM25 TF weighting rewards documents with many, many instances of a term in them. Explain why that's the case. (2)

Question 6

Topic: Lecture 8

Source: Lecture 8

What are some assumptions that we make when we are interpolating between a document and a corpus? When should we trust the corpus more, and when should we trust the document more? (2)

Question 7

Topic: Lecture 5

Source: Lecture 5

What impact do sparse matrices have on similarity metrics like cosine similarity? (1)

Question 8

Topic: Lecture 6

Source: Lecture 6

Why don't we just use k-means to cluster document-vectors (sparse or dense)? (1)

Question 9

Topic: Coding

Source: Coding

Write a function that returns the most likely n documents given a term-document matrix, a smoothing parameter, and a query. (3)

END OF QUIZ