

START OF QUIZ

Student ID:

35366434, Choudhary, Tushar

Question 1

Topic: Lecture 7

Source: Lecture 7

Explain why boolean filtering is usually insufficient for retrieval, and why we normally need some way of scoring the documents. (2)

Question 2

Topic: Lecture 6

Source: Lecture 6

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

Question 3

Topic: Lecture 7

Source: Lecture 7

What is the benefit (in terms of efficiency) of placing the most discriminative search terms first in a boolean search? (1)

Question 4

Topic: Lecture 5

Source: Lecture 5

Why can we be confident that a low-rank approximation of a matrix contains the most important information in a document? (1)

Question 5

Topic: Lecture 5

Source: Lecture 5

Why can we represent a rank- m matrix as the sum of m rank-1 matrices *or* the product of an $n \times m$ matrix and an $m \times n$ matrix (ie, what is matrix multiplication doing that we can take advantage of)? Explain. (2)

Question 6

Topic: Lecture 8

Source: Lecture 8

Why don't we use a higher-order language model to perform IR? (1)

Question 7

Topic: Lecture 6

Source: Lecture 6

In some ways, we could consider Beta distributions themselves to be an embedding of a topic. Explain, and explain how we might be able to leverage that. (2)

Question 8

Topic: Lecture 8

Source: Lecture 8

What do we mean by interpolation? (1)

Question 9

Topic: Long

Source: Lecture 5

Imagine that we are working with a language other than English, such as Indonesian, with significant agglutinative morphology (words are inflected through the concatenation of affixes to a lemma). How do you think that this would impact our various vector space models? Which of them would be most affected, and which would be least affected? Explain. (3)

END OF QUIZ