

START OF QUIZ

Student ID:

**97233886,Nandaku-
mar,Hariharavarshan**

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 7

Source: Lecture 7

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

Question 3

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 4

Topic: Lecture 6

Source: Lecture 6

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

Question 5

Topic: Lecture 8

Source: Lecture 8

What do we mean by interpolation? (1)

Question 6

Topic: Lecture 5

Source: Lecture 5

Explain the logic behind the IDF part of TF-IDF (ie, why does it give higher weights to more "interesting" words?). (1)

Question 7

Topic: Lecture 8

Source: Lecture 8

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