START OF QUIZ
Student ID:
74832403,Kumar,Rakesh

Question 1

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 2

Topic: Lecture 7
Source: Lecture 7

What is the purpose of an inverted index? (1)

Question 3

Topic: Lecture 6
Source: Lecture 6

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

Question 4

Topic: Lecture 5
Source: Lecture 5

What advantages do sparse vectors have over dense ones. (1)

Question 5

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 6

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

Question 7

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 x m matrix and an m x n matrix (ie, what is matrix multiplication doing that we can take advantage of?)? Explain. (2)

Question 8

Topic: Lecture 8
Source: Lecture 8

What is the reasonining behind substituting TF-IDF with Okapi BM25? (1)

Question 9

Topic: Coding
Source: Coding

Imagine that our corpus contains 1M documents. We have 3 queries that we are looking at. The first query has 5 relevant documents, returned in positions 1, 5, 10, 20, and 50. The second query has 3 relevant documents, returned at 10, 11, and 12. The third query has only one relevant document, and it is returned in position 7. What is the MAP of our system? (3)

END OF QUIZ