START OF QUIZ
Student ID:
22039382,Amal,Cenith

Question 1

Topic: Lecture 8
Source: Lecture 8

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 2

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 3

Topic: Lecture 5
Source: Lecture 5

We often weight our matrices using something like PMI or TF-IDF. Do you think it would make sense to do this after applying SVD? Why or why not? (2)

Question 4

Topic: Lecture 8
Source: Lecture 8

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

Question 5

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

Question 6

Topic: Lecture 7
Source: Lecture 7

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

Question 7

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

Question 8

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

Question 9

Topic: Long
Source: Lecture 7

Imagine that we have 2 information retrieval systems, and we are evaluating on the same test set, which has 10 relevant documents. The first system returns them in positions [1, 5, 7, 15, 25, 50, 60, 70, 71, 90]. The second returns the documents at positions [2, 3, 6, 8, 10, 62, 80, 83, 91, 95]. Make an argument for each system being better, and provide support for both. Explain which system you would rather use, and why. If there are any other considerations, list them. (3)

END OF QUIZ