START OF QUIZ
Student ID:
52196086,Ahluwalia,Max

Question 1

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 2

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 3

What is the benefit of evaluating boolean queries using set operations instead of loops? (1)

Question 4

Topic: Lecture 8
Source: Lecture 8

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

Question 5

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

Question 6

Topic: Lecture 6
Source: Lecture 6

In class, we saw a few topics that we were unable to identify. What could be a cause for such pointless topics (ie, how might we ensure that our topics are better? (2 reasons). (1)

Question 7

Topic: Lecture 6
Source: Lecture 6

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

Question 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 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