

START OF QUIZ
Student ID:
37083607,Zeng,Jiao

Question 1

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 2

Topic: Lecture 5

Source: Lecture 5

Why do we need methods like t-SNE? (1)

Question 3

Topic: Lecture 7

Source: Lecture 7

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

Question 4

Topic: Lecture 6

Source: Lecture 6

Why can't we just run an HMM over documents to discover the latent states like we do for POS-tagging? (1)

Question 5

Topic: Lecture 8

Source: Lecture 8

What do we mean by interpolation? (1)

Question 6

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 7

Topic: Lecture 8

Source: Lecture 8

In class (and in the lab) you saw some examples of using a language model for IR. How do you think we could incorporate an LLM into the IR pipeline? In what ways do you think an n-gram lm might be more appropriate? (2)

Question 8

Topic: Lecture 7

Source: Lecture 7

Explain why the cosine similarity between a (TF-IDF-weighted) document and query vector is roughly equivalent to adding up the TF-IDF scores of each word in the document that occurs in the query. (2)

Question 9

Topic: Long

Source: Lecture 8

In class, I mentioned that we might want to provide a weighted ranking of documents in an IR system (for example, we might have a list of relevant documents already sorted for relevancy, and we want our IR system to not only return those documents high in the returned document list, but in the same order). How might we modify the MAP metric to reward the ordering of the documents, as well? Explain (use an example, if you have to). (3)

END OF QUIZ