



UNIVERSITY OF SOUTHERN MAINE

Text Mining and Analytics, Spring 2024, Assignment 5

Instructor: Behrooz Mansouri (behrooz.mansouri@maine.edu)

Due: April 24, 2024

This is a reading and application assignment to review concepts related to information retrieval. Carefully read the paper "[ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT](#)". Then provide two paragraphs summarizing your understanding of the approach. Then provide a third paragraph, discussing how this model is different from the Bi-encoder and Cross-encoder models from Lecture 20.

In the next step, index all the songs from the training and validation sets you have created for assignments 3 and 4. Then use songs from the test set as the input query and for each query, find the top-10 relevant results.

Finally, provide a sample code on how you can fine-tune the ColBERT model. Your code should specify the input format with small sample files. You do not need to search with a fine-tuned model.

Deliverables:

Task 1: .pdf file having your answer

Task 2:

- .py file for running your code
- Retrieval results file in the Trec format, with song names as the document ID
- README file providing detailed instructions to run your code

Task 3:

- .py files for fine-tuning ColBERT model
- README file providing details on how to run the code, and input file formats
- Small sample files for running your code

Grading:

This assignment is done individually and has 5 points.