



# UNIVERSITY OF SOUTHERN MAINE

## Introduction to Information Retrieval, Fall 2024, Assignment 4

**Instructor: Behrooz Mansouri**

**Due: November 12, 2024**

---

In this assignment, you will explore the application of large language models for improving search results. The goal is to find relevant answers to given questions about traveling. The data files remain similar to the previous assignments.

### Applying Large Language Model

We have visited several ideas and approaches to use large language models in information retrieval. In this assignment, each student will pick an approach and study it for information retrieval. You should carefully explore your options for the language model that you pick and the technique you apply.

After you pick your LLM and technique, you should try two different prompts and in your report discuss which one was better and why.

#### Important notes:

1. If you are using the CS servers, please make sure to use netstore to save HuggingFace models on netstore, following the command shown in slides. Overusing the server space can and will suspend your account from CS machines. And please remove your models later.
2. As LLMs can be time consuming, it is recommended to select a set of candidates using models from the previous assignments and re-rank their results.

### Evaluation Measures

Evaluate your models reporting the  $P@1$ ,  $P@5$ ,  $nDCG@5$ , MRR, MAP. For metrics other than  $nDCG$ , the relevant documents are those with scores 1 and 2 (in the qrel file). You can use any tool that you want for evaluation. However, this tool should be standard, and the results should be reproducible (Trec\_eval and Ranx are recommended). For each topic, you should exactly return 100 results.

For your report, you need to include a table as follows for the topics in topic 1 file:

Model	P@1	P@5	nDCG@5	MRR	MAP
Prompt 1					
Prompt 2					

## Report and Analysis

In your report, you will explain your technique, followed by prompts that you have used. (If you are using system and user messages you should split them and clearly indicate them in your report.)

You have two models (prompts) that should be used on both topic\_1 and topic\_2 files. Make sure the result files can be passed to Trec\_eval and Ranx for evaluation using the correct format. You do not need to provide an evaluation script. Complete the table above (averaged over all the topics). Then provide a ski-jump plot based on P@5. Using this plot, discuss and reason about the success and failure of your system. You should provide at least one pair of successful and failed (topic, answer) in your discussion for each model. You should provide topics, and retrieved instances to support your claims (using the text not IDs).

### Notes for submission:

1. Python file/files (only .py is acceptable, not .ipynb) with codes for all retrieval. Codes should be well-structured with comments to run. You should have a README file that provides clear guidance on how to run your code and get the result files. **Explicit file paths should not be used in the code**, and topic files should be passed as arguments to your code.
  2. You will get 4 retrieval result files by the end of this assignment. You should submit them all following the naming convention below:
    - prompt1\_1.tsv: prompt 1 on topic\_1 file
    - prompt2\_1.tsv: prompt 2 on topic\_1 file
    - prompt1\_2.tsv: prompt 1 on topic\_2 file
    - prompt2\_2.tsv: prompt 2 on topic\_2 file
  3. A .pdf file with your results table, ski-jump plot, and analysis. In this file, you will also explain what model you have chosen and justify your choice.
  4. Any assumptions made by students should be explicitly mentioned in the submitted
- Note:** Any submission not in the format explained above will be dropped, resulting in 0, without the possibility of regrading