LiveDataLab System Extension - Project Proposal

We have chosen to do the system extension - specifically related to improving LiveDataLab. This project will focus on linking relevant lectures to MPs in LiveDataLab to provide a resource for students who would like additional background information for MPs. We will be using the BM25 ranking function to determine relevant lectures, where the given search query is the MP description and the documents to be ranked are the lectures’ transcripts. The MP page will display the 3 most relevant lectures by default, but the user will be able to change the number of relevant lectures they would like to see. In order to complete this project, the programming language we have chosen to use is python.

The datasets we will be using in our project will be constructed from the transcriptions in Coursera. We will also be using the README files in GitHub from the MPs in the class. From the transcriptions, we will find the keywords in the lectures and rank them based on their importance. The main algorithms we will be using in our project are the search function, to search words correlating to lectures in GitHub, and a ranking function, which will rank the importance of the words in lectures and how relevant they are to the topics the lecture talks about. We will be assigning weights to key terms in each lecture and if they are mentioned in any of the MPs, we will link them to allow for easier access for students working on the MPs.

Currently, there is no feature that displays relevant lectures for the MPs on LiveDataLab. Once we add the function, the 4 of us will use relevance feedback to check each recommended lecture video for each MP and determine if it is relevant and if the lectures contain information relevant to us when we did the MPs ourselves. We will also check to make sure the lectures are displayed in order of relevance from most to least relevant.

Our code is meant to be an extension of the LiveDataLab that will automatically link MPs with relevant lectures. As a result, it will utilize LiveDataLab’s access to the GitHub MP repositories to obtain the relevant keywords. Furthermore, it will enhance the system by providing an option for students to gather the relevant lectures for the MP based on the query.

The proposed project has an estimated workload of roughly 80 hours (20 \* 4). The project is split up below into a weekly schedule that will take us to the end of the semester. We must first become familiar with the existing system, obtain the query, create the function, implement the ranking function, create the UI display, test every feature, and create the presentation over the course of the next 6 weeks.

* Oct 31 - Nov 4: Get familiar with the LiveDataLab system, create our function, and get search query(MP description)
* Nov 7 - Nov 11: implement BM25 ranking function
* Nov 14 - Nov 18: progress report, get documents(lecture transcripts), create UI display
* Nov 21 - Nov 25: testing and additional features
* Nov 28 - Dec 2: work on final documentation, submit code
* Dec 5 - Dec 8: final presentation