Text Information Systems
Project Proposal

Content:

1. Name and netid:

Name: Nita Agarwal
NETID: nitaa2
TeamName: TeamNA (individual participant)

2. Theme of the Project

The Theme of my project is Free Topic

3. Free Topic details:

Develop an application to link the MP descriptions to relevant TIS course lectures provided on Coursera. This application will help the learner to work on a MP effectively by suggesting any relevant background lectures, if needed. I plan to create multiple components to successfully implement this application. First, collect & preprocess all the MP assignment descriptions. Second, collect & preprocess all the lecture video transcripts. Index these lecture transcripts so that they can be searched for corresponding MP assignments. Third, come up with the ranking function to rank the lectures based on the MP description. Fourth, evaluate the ranking and fine tune the parameter settings based on the human evaluation of the ranked list.

The dataset involved is the MP assignments provided in the course (MP1, MP 2.1, MP 2.2, MP 2.3, MP 2.4 & MP3) & the transcripts for the lecture videos from week1 – week 12. I plan to use Python, metapy library for the implementation. The evaluation of the effectiveness of the application can be done based on the how well the relevant lectures are suggested by the application corresponding to a given MP description.

4. Programming Language:

I plan to use Python as the programming language for implementation of my project

5. Task list for the Project:

I plan to work on the below task list for the implementation of the application

- **Task A: Collecting data (5 hours):** Gather MP assignment and lecture transcript from Coursera for the complete course duration

- **Task B: Preprocess data (8 hours):** Data preprocessing may include removing stopwords and words with less than min length, stemming etc. This step will help in creating the corpus and query data set.

- **Task C: Build Ranking logic (5 hours):** Build the logic for indexing the corpus data set and determine the appropriate metapy ranker for the application

- **Task D: Evaluation and parameter fine tuning (5 hours):** Evaluate the ranking and fine tune the parameter settings based on the human evaluation of the ranked list**.**

6. Stretch goals for the Project:

If time permits, I would like to work on the below stretch goals

Task E: Build User Interface for the project. This will help the user select the MP and see the corresponding lecture video links on a UI instead of command prompt.

Task F: Build Web based API driven interface for the project. This will help in accessing the application over web