**CS410 Project Proposal**

**Group “TIS Chicago”**

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

Josephine Falso, Team Captain, jfalso3

Sree Alaparthi, [sa68](mailto:sa68@illinois.edu)

Narasimha Kethireddy, nrk6

The GitHub project URL is as follows:

https://github.com/JosephineFalso/CourseProject

1. What topic have you chosen? Why is it a problem? How does it relate to the theme and to the class?

We have chosen a project within the topic of “Intelligent Learning Platform”. We will build a back-end application to identify and explain popular keywords in the Coursera lecture videos for CS410. Our application will provide a cohesive integration of the instructor’s lectures with the student discussions in CampusWire.

Significant keywords will be highlighted in the transcription text. When the learner hovers over a keyword, the hover text will provide some additional information from the CampusWire posts related to that keyword.

This alleviates the inconvenience of navigating in and out of Coursera. By incorporating all of the relevant information directly in Coursera, the learner will have a more focused and productive learning experience. The techniques applied to this project demonstrate mastery of both text access and text mining.

1. Briefly describe any datasets, algorithms or techniques you plan to use.

To build this application, we will perform web scraping of the transcription text in the Coursera CS410 lecture videos. We will build a simple website to store this scraped text. We will use a Python NLP library such as NLTK, PKE, or spaCy for text mining. We will also perform topic analysis and keyword identification via a search of CampusWire postings to identify and associate the postings to the appropriate keywords from Coursera.

1. How will you demonstrate that your approach will work as expected? Which programming language do you plan to use?

This project will be developed in Python.

We will demonstrate the approach by randomly selecting keywords and verifying that the information provided in the pop-up window matches the number of true references in CampusWire. We will limit the scope of the project to a subset of the CS410 Fall semester content. The subset will be a selection of weeks of material or a selection of course videos depending on the time that we have.

1. Please justify that the workload of your topic is at least 20\*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

The TIS Chicago team consists of three members. The workload of sixty hours will be distributed as follows:

Project Proposal, 1 hour

Progress Report, 2 hours

Software Code Submission, 60 hours as follows:

* Web Scraping of transcription text in course videos of CS410 lectures and Text Mining of keywords, 20 hours
* Search CampusWire feed to associate posts to keywords, 20 hours
* When user hovers on keywords, display CampusWire information, 20 hours

Documentation, 2 hours

Software Tutorial Usage Demonstration, 2 hours