

CS 410 Final Project Proposal October 2021

Topic: Topic Modeling on Top Topics Discussed in Quora on Illinois CS Department

Team: Panda5

Team Captain: Sushanta Panda

Sushanta Panda **NetId:** Panda5

1. Introduction:

For the final project, I choose the topic of Topic modeling from the Quora feed. I will be extracted all the Quora feed which talks about the Illinois Computer science department and stored it in a text file. This will be a manual feed and all the text will be stored as plain text. The system will then be built over python to do the Topic Modeling of identifying the Top 10 Topics.

2. Overview:

The Quora is the platform where the questions and answer are being made by users. This platform was a great platform for the users to raised questions and many users responds to it based on their prior experience and will be very useful to the end user who raised the question. As part of the Project, I will be doing the Topic modeling to fetch out the Top Topics discussed in Quora about the Illinois University for Computer Science (CS) department. First, I will fetch out (manually) the Quora feed about both the questions and answers about the Illinois University's Computer Science Department and stored in the Text file. Then will have a program in Python which will do Topic modeling and fetch out the Top Topics discussed by users in the Quora. This will be present as the form of list from the Program, where I can show the Top 10 or 20 lists based on the limit set in the Program.

3. Motivation:

Student often query and research about many universities before they finally apply for the university. This Topic modeling will be helpful to understand what people discussed in Quora about the Illinois University for the Computer Science Department. If let's say the Topics comes out as "Research" / "Project" / "Student" / "Professor" would give more confident to a student rather "Hectic" / "Cost" / "Simple". This also gives an uber picture of the department what people should discuss about. Also this program can leverage to other Question & Answer Site (or social media platform) or to other university tool

4. Work To Be Done:

I need to figure it out how to do the Topic modeling in Python after the cleaning / pruning / tokenization of the Quora Feed. Also needs to research the Quora feed to see enough discussing available for the Computer Science Department for the Illinois University.

5. Evaluation:

I will be manually checking all the Quora feed and do some research in the Google to verify whether the Top Topics pulled out by my program is matching and in-line with the Topic generally people are discussing.

6. Tools:

- Python 3.5
- NLTK Natural Language Tool Kit
- Numpy
- Pandas
- Scikit – Learn, Sci-py