

CS410 Project Proposal

Dancing Text: Yuxiang Liu (yuxiang@illinois.edu, leader), Hongfei Ma (hongfei7@illinois.edu)

1. Proposed Study

We plan to improve the **EducationalWeb System** through automatically creating teaching materials for in-demand skills, which is an extended version of the existing system. To satisfy the increasing demands from people who are looking for high-quality education but cannot get access to it whenever necessary, we aim to **identify in-demand skills** and **create lectures and tutorials for these skills**.

To identify in-demand skills, we will crawl data from social media and identify emerging keywords or topics through topic extraction. To create lectures and tutorials for skills, we will recommend most relevant slides given specific topics. In this project, we plan to mainly use python and javascript languages.

2. Impacts Of The Proposed Work

Through this project, we would enable emerging topic identification and slides recommendation for topics, which will help users who are interested in getting high-quality education materials at any time.

3. Project Description

3.1 Identifying in-demand skills

We plan to crawl tweets from Twitter. A hashtag is a type of tag that people usually treat as the theme of the corresponding tweet, so we would directly consider hashtags as the topic of tweets. Then, we would use features of tweets such as retweets, replies, or likes to identify which hashtags are emerging or popular keywords/topics. In particular, we may limit our topics within computer science or STEM disciplines to reduce/avoid skill-irrelevant topics. After we obtain these skill-related topics, we would use them in the next part of creating corresponding lectures and tutorials.

3.2 Creating lectures and tutorials for skills

In this part, we will first crawl slides of other courses in UIUC. For these slides, we will identify the topics and generate training data. Next, we will propose an efficient ranking algorithm to automatically select relevant slides for each specific topic. Once a user selects a topic from those obtained in the first part, the system will present relevant slides to the user on the webpage. To demonstrate the performance of our algorithm, we will separate our crawled slides, together with existing slides, into positive samples and negative samples, and compare our algorithm with classical algorithms such as BM25 with this generated dataset.

4. Timeline

Task	Estimated Time Cost	Expected Completion Time
Social media crawling	5h	Week 11
Paper survey	5h	Week 11
Emerging topic identification	15h	Week 14
Slides crawling	5h	Week 11
Rank algorithm design	15h	Week 13
Integration with existing system	20h	Week 15