

CS410 Final Project Progress Report

Project Topic: Intelligent Learning Platform: Organize the scattered lectures into a coherent “multimedia textbook” and create an index

Group Name: Freelancer

Member: Jiecheng Zhao (jz109@illinois.edu)

1. What tasks have been completed?

At this time, I have completed two tasks: literature review and keywords extraction function development.

In the literature review, I have reviewed books, papers and guidance about text indexing techniques, text coherence, and ranking. I have also taken a deeper dive into the MeTA toolkit and FLASK.

In the keywords extraction function, the contents are extracted from the corpus and the timestamps are discarded. MeTA is used to tokenize the contents, exclude the stop words, and get the highest occurrence of the candidate terms.

2. Which tasks are pending?

The pending tasks includes:

- 1) Keywords and topic indexing function development
- 2) Webpage output development, and
- 3) Test and debug.

3. Are you facing any challenges?

Yes. One of the major challenges I have experienced at this stage is to refine the candidate terms.

- 1) Some terms may stem from the same source. A stemmer needs to be used as the filter.
- 2) Some terms have high occurrence and are not in the stop word dictionary, but they are not quite the keywords of the course. For example, in this text information system course, compute appears many times, but does not qualify for a keyword of this specific course. One possible method is to get the background language model of general computer documents, and filter out these terms.