Text Miners Team Project:

* Progress made thus far:

Scale up the current system:

POC to read the pdf file(slides) from UIUC url and download it in the required format is complete.

Implemented the changes to show the newly added courses for users to select on test data.

* Remaining tasks:

Creation of Automatic Web Crawler by leveraging the POC that we completed now.

Implementing the code changes for search result and recently visited slides.

Unit Testing and Integration Testing

* Any challenges/issues being faced:

Determining the reference link on Course era was not straight forward as the links contained sections which could not be parsed via code. We need the slides in single-page pdf as input to the educational web search program. Also, it requires additional authentication, ultimately downloading the pdf from cloudfront.net. As we don’t have access to cloudfront, we decided to refer slides from UIUC site as input data to our program.