Team: Early Birds (Progress Report)

1) Which tasks have been completed?

- Initial data wrangling (corpus comprised of full transcripts without regard to timestamps)

- Preliminary search functionality (BM25/Okapi, untuned)

- Creation of test query files for our search functionality

- Integration of search engine into backend API

- Deployment of backend API to cloud service

- Deployment of front-end component to cloud service, and integration with the back-end API

- Mock-up of front-end website (with media player)

- Upload course content including video, text and pdf files to AWS S3 bucket

- Architecture diagrams and instructions for building the project and running locally added to the Github repository README: https://github.com/tghabour/CourseProject

- Initial prototype: https://stingray-app-3sdu7.ondigitalocean.app/

2) Which tasks are pending?

- Update the search functionality to return correct text, video and pdf AWS S3 links.

- Add next, previous buttons to video player for user to navigate through search results

- Finalize additional corpus with timestamp information to enable more granular/refined search results.

- Use scene detection on times to add topic parameters to transcripts.

- Ranking function evaluation and optimization.

- Instructions and documentation

- Tutorial video

3) Are you facing any challenges?

- Inverted index and ranking functions were implemented with metapy, which requires Python version 3.5. Ideally, however, we were hoping to use Django REST to create the API, which requires version Python 3.9.

  - We were unable to use a fork of metapy that supports python 3.9 because it failed to compile on MacOS. We resorted to using python 3.5 and an earlier version of Django, but spent some time trying to use python 3.9.