

# Project Report: Hierarchical Lecture Summarizer

**Team Name:**

Group of One

**Team Members:**

Alexander Ho (ajho2) (Captain)

**GitHub Link:**

<https://github.com/Alx-Ho/CourseProject/>

**Topic Chosen:**

Hierarchical Lecture Summarizer (Intelligent Learning Platform track)

1. Which tasks have been completed?

I have completed a preliminary version of the hierarchical clustering based on a custom loss function which is then optimized by the Python package **hyperopt** which does an informed hyperparameter space search. It has been tested on some preliminary data, and takes into account the average Jaccard similarity across documents within the same cluster based on top k n-grams, the mean cosine distance between cluster centroids, and the number of clusters that contain few samples or too many samples. I have also downloaded much of the lecture content for processing and chunking for input into the clustering model.

2. Which tasks are pending?

Currently, I am building the final dataset based on the lecture content as well as building out the interface with which the users will interact with this hierarchical clustering algorithm and summarizer.

3. Are you facing any challenges?

The primary challenge at this point is creating the interface for interacting with the hierarchical summarizer. There are a few different approaches that I have experimented with, but the final form has yet to be finalized. Designing this interface will take the majority of the remaining work.