

Team KAF4 Members: Captain Kristen Fisher [kaf4]

Topic: Intelligent Learning Platform: ConceptView

Goal: Extract concepts taught from lectures in to build a “concept summary” for each lecture.

If concepts from lecture materials can be extracted intelligently, users can efficiently retrieve lectures containing relevant information based on these generated concept summaries.

### **Progress Report**

At this point, I have extracted the text data needed for this project. This includes both transcripts as well as the text on the slides from all the lectures available on Coursera. I have cleaned the data to remove delimiters and other strange characters to create usable datasets. With this data I have been testing the extraction of n-grams and building a robust list of stop words. For evaluation I am using the provided ConceptView data to ensure the terms I have been extracting seem to include appropriate concepts covered in the lecture. Initially, I was only planning to use the transcripts, but I found including text from the slides seems to improve results in my testing.

This week I plan to spend time researching additional topic modeling tools and techniques I can incorporate beyond the scope of the basic stuff I have been doing with MeTA during dataset testing. After that, the next stages will be testing out some of the tools I am researching, figuring out a way to build the concepts after filtering and/or ranking the terms, and then lastly generating the final summary from those results. I have not faced any significant challenges yet, but I think rebuilding a coherent summary from basically a bag of words might be difficult with limited human effort. However, this is something I plan to spend some time researching as well.