

1) Which tasks have been completed?

We have mostly completed task 1) Web scraping data/course from Coursera. We used beautiful soup to pull the html files, grab links and headers and parse to create a list of all the Coursera pages (including videos). As well, we parse through the pages to grab the actual document from the pages.

We are also working on task 2) Search Engine Algorithm, using the documents (reading pages only currently) and using base bm25 (changeable as we test other algorithms) as our search engine algorithm.

2) Which tasks are pending?

We have not started on 3) Video Search Algorithm or 4) Recommender based Algorithm.

For 3) we are planning on similarly using a base bm25 model to parse through the video contents (since the transcript is available).

We have not really thought much about 4) yet.

3) Are you facing any challenges?

One of the biggest challenges is actually implementing the algorithms and connecting it with an interface which we will have to use later on in our final product