

Which tasks have been completed?

I have completed the task of creating a dataset for the paper recommendation system. To do this, I utilized the Arxiv API to obtain the HTML formatted paper responses. I then parse these responses using beautiful soup and regex libraries. My current dataset contains 8 topics within the computer science field with about 100 papers for each. I've also started researching which model to use and am working on my implementation on Pytorch. I am currently using a transformer encoder architecture with a linear classifier and training the entire model from top to bottom.

Which tasks are pending?

Tasks that are still pending include keyword matching, writing the API, and evaluating the classification system.

Are you facing any challenges?

I'm currently trying to debug some issues with padding and masking since the inputs can be of various lengths with batching. I am also noticing that the loss is not converging smoothly when running it with a batch size of 1.