So my first week at Fandor has been pretty great. Everyone on the team is nice and super helpful. The first couple of days were a bit of an information overload, but then I started to get familiar with the database schemas and I started poking around into allllll the data the company records (a LOT of different things wow).

The project I’m working on this summer turned out to be a combination of my two proposed projects, which is pretty great because I didn’t want to choose between them. I’ll be building a recommendation engine for editorial content, starting simply with a few recommended articles at the end of each editorial on the site. “Starting simply” is still pretty complicated. It turns out there’s a lot that goes into planning what sort of data you want to use, which metrics make the most sense to use and what we actually care about measuring, and how the engine will work with unfamiliar data. Lots of fun ahead!

I’m working, now, with an elasticsearch structure as a data storage system. Each post will have different entities related to it, with some possible other metadata as well, stored in json format. Today I started thinking about the storage and metadata structure that makes the most sense to relate thousands of documents together using over 50,000 unique entities (films, directors, concepts) and it is not a trivial task to plan for the data you might want to use in the future.

What I wish I had more experience with right now is SQL queries. They are the backbone of the analytics at Fandor and I’m glad to be learning just how powerful they are and what sorts of questions they can answer. More SQL!