Using NLP to Study Steven King's Dark Tower Series
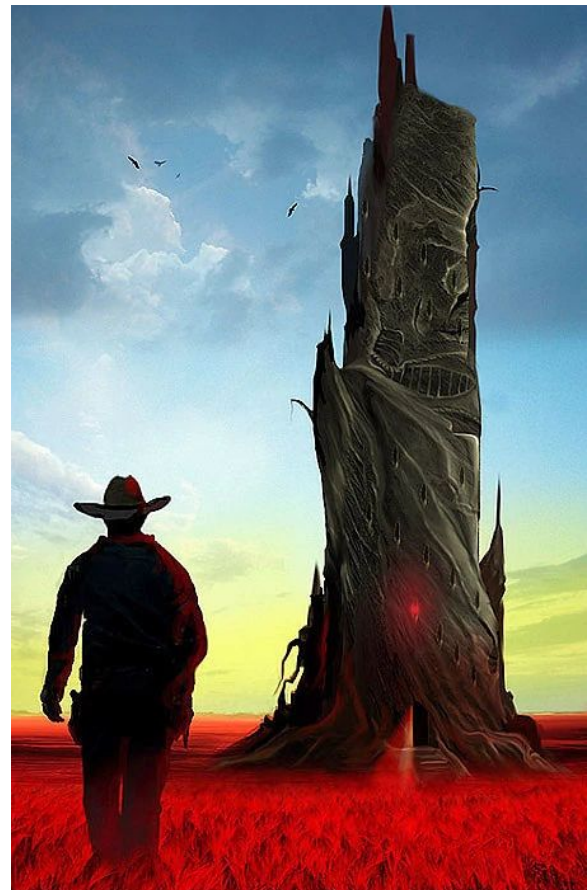
Derek Nelson

Our journey to the Dark Tower

The Dark Tower is very *meta*. In the series it is said that "everything leads to the Dark Tower" and that is mirrored in the writing of the books themselves.

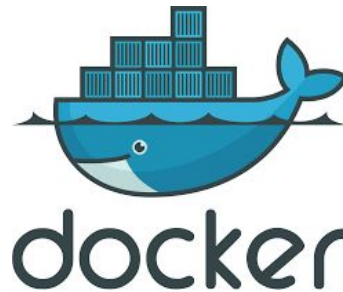Many characters especially from King's early work ended up in the Dark Tower multiverse.

I wanted to use NLP to map the relationship between the series and his second book *'Salem's Lot*.

Tech

- Mostly Java with Spring
- Gradle and Gulp for builds
- Neo4J for Storage
- D3 for Visualization
- OpenNLP and CoreNLP for text processing
- Docker for deployment
- Calibre for converting EPUBs to plaintext

Starting NLP

I started with **Named Entity Resolution** (NER) using books from **Project Gutenberg**. I went pretty fast but I quickly realized I needed much better data. From here I tried **Co-Reference Resolution** to map pronouns to characters.

I had memory problems and it took me multiple hours finish a single book. When I was done all of the entities and references were wrong.

Eventually I learned to configure CoreNLP and was able process books in under 5 minutes. **David Bamman** from *UC Berkeley* saved my butt.
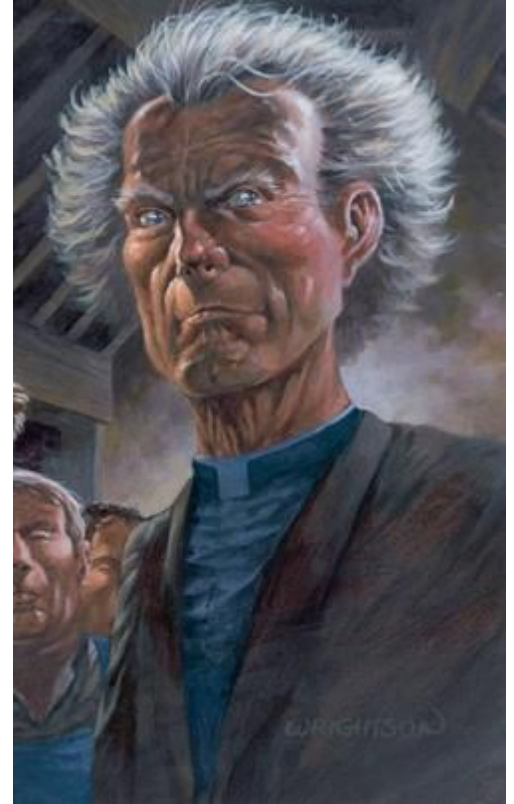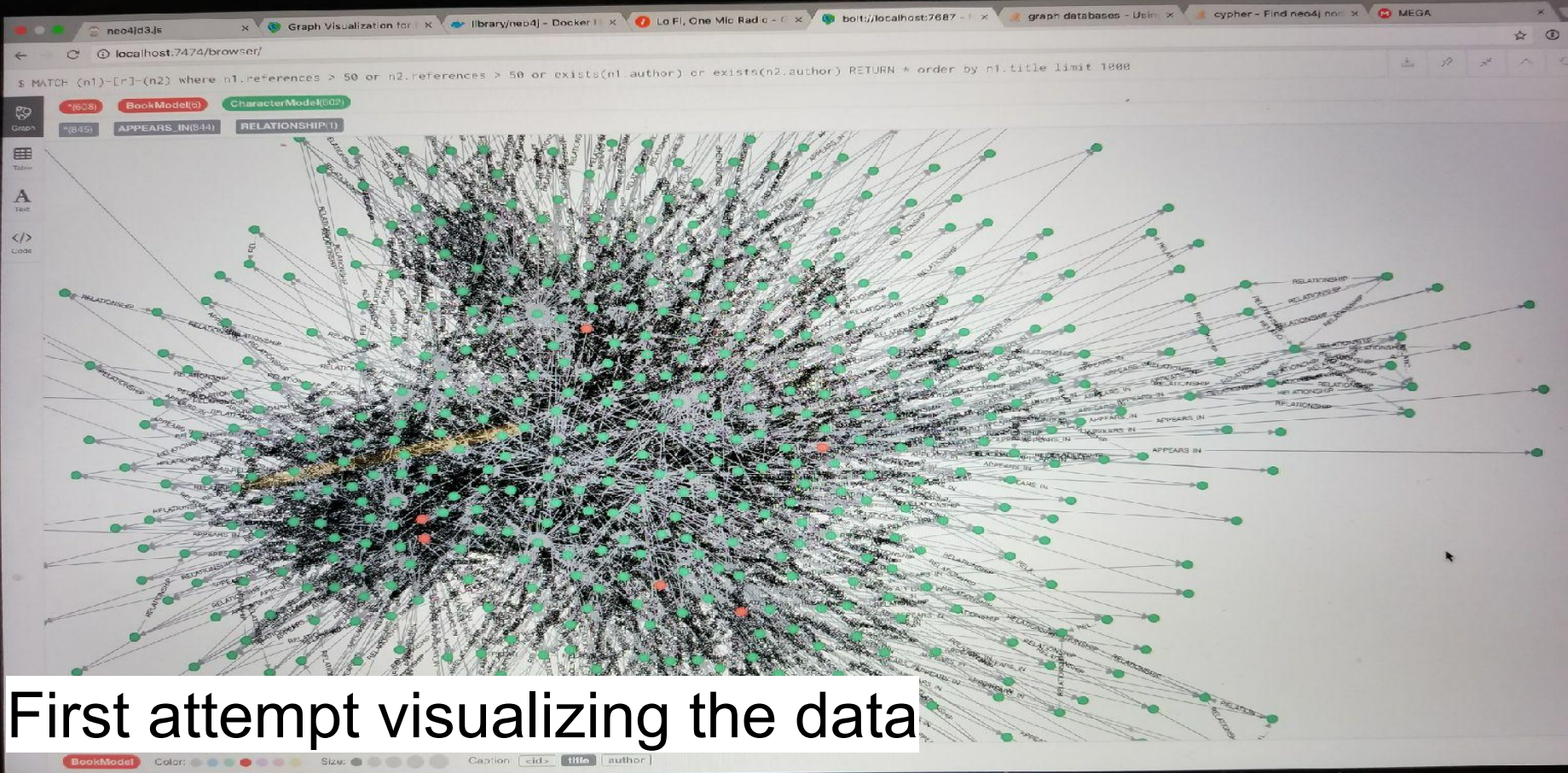

"I voted for Nader because he was most aligned with my values," she said.

Things were good for a bit

I quickly realized I would need to use Spring and a build process more robust than Ant though. Mr. Bammam's code wasn't compatible so Saturday was mostly rewriting everything.

By 3am on Sunday I had processed and saved all of the Dark tower books as well as 'Salem's Lot.
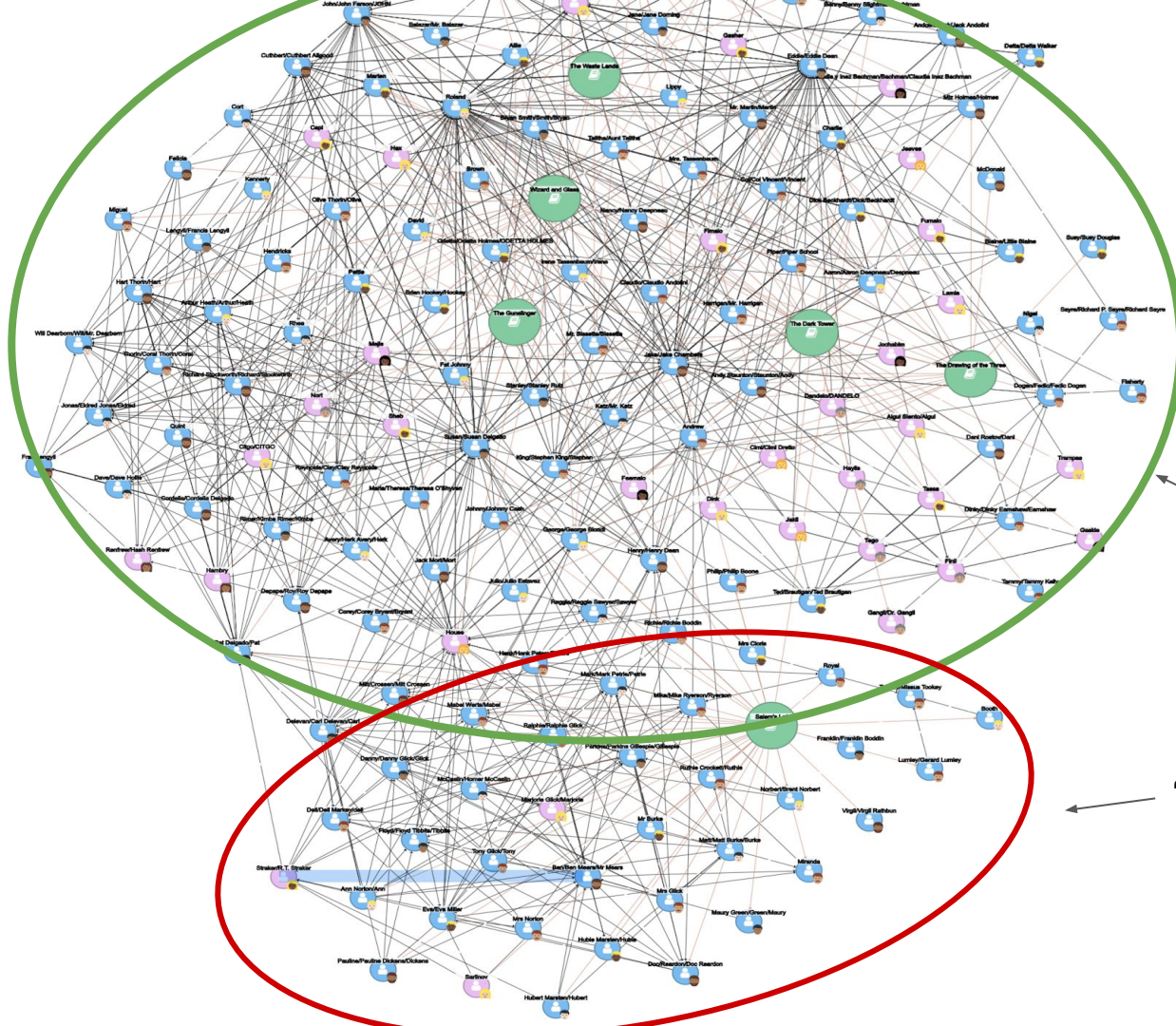
First attempt visualizing the data

Visualization

Sunday was spent cleaning data and **HEAVILY** modifying Neo4J's built in renderer (which just happened to be written in D3). Mostly displaying specific metadata and teaching the graph how to "organize" itself to best show the relationships.

Success!

Dark Tower

'Salem's Lot