

Using NLP to Study Stephen King's Dark Tower Series

Derek Nelson

Imes/ODETTA HOLMES

Irene Tassenbaum/Irene

Claudio/Claudio Andolini

Hamigan/Mr. Hamigan

Piper/Piper School

Aaron/Aaron Deepneau/Deepneau

Jochabim

Andy Staunton/Staunton/Andy

Jake/Jake Chambers

Mr. Bisette/Bisette

The Gunslinger

Fat Johnny

Mejls

Brian Hookey/Hookey

Pettie

If at his counsel I should turn aside
Into that ominous tract which, all agree,
Hides the Dark Tower. Yet acquiescingly
I did turn as he pointed: neither pride
Nor hope rekindling at the end descried,
So much as gladness that some end might be.

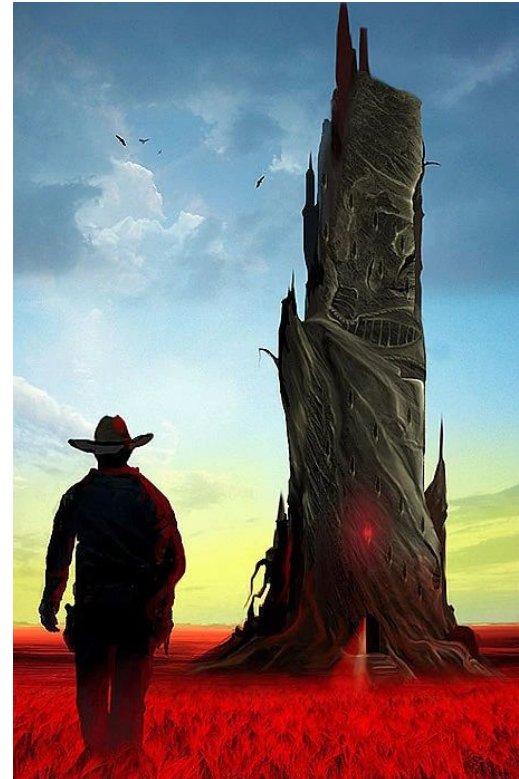
- Robert Browning

The Dark Tower is very *meta*. In the series it is said that “everything leads to the Dark Tower”. Many characters especially from King’s early work ended up in the Dark Tower multiverse.

In the Dark Tower Ka is is a force, a guide, and a path.

Ka-tet is a group of people pulled together by Ka.

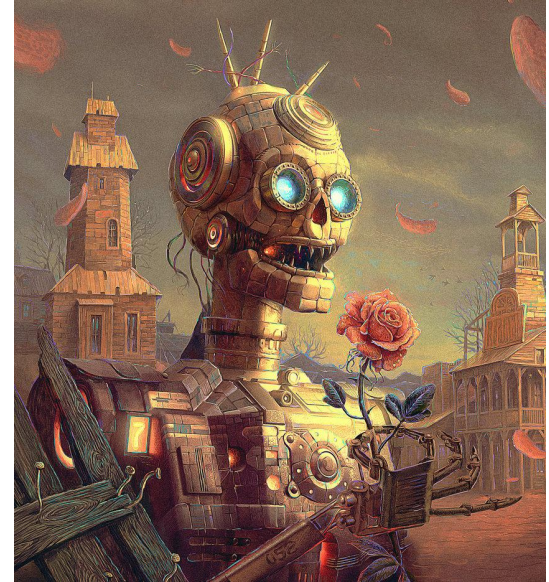
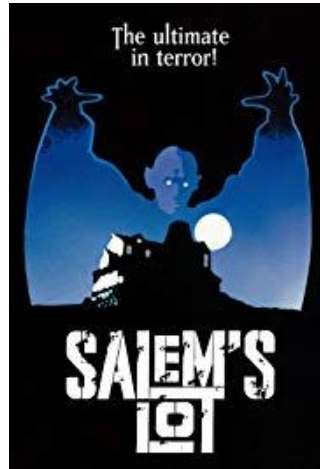
We are ka-tet. We are one from many



Goals

My goal is to analyze The Dark Tower and its relationship to King's second book '*Salem's Lot*'.

Like Roland we will follow the Beam, explore the multiverse, and discover the "Dark Tower."



Terms

Noun phrase (NP) - a collection of words which references persons, places, or things.

Animate - A feature of some languages which expresses sentence of a **noun phrase**. *Associates, attendee, adventists.*

Pronouns - A word that substitutes for a noun phrase.
They/them

Named Entity Resolution - Flags certain nouns as “names.” *Niagra Falls* is a named entity.

Co-Reference Resolution - to map pronouns to entities.



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

Terms continued

Anaphor - a phrase whose meaning depends on another expression.

Cataphor - a specific anaphor where the meaning is after the anaphor.



Problems

Split antecedents

“Lisa smiled at Bart . The children were happy..” [\[Source\]](#)

Cataphora

*“From the corner of the divan of Persian saddle-bags on which **he** was lying, smoking, as was **his** custom, innumerable cigarettes, **Lord Henry Wotton** could just catch the gleam of the honey-sweet and honey-coloured blossoms of a laburnum”* - Oscar Wilde

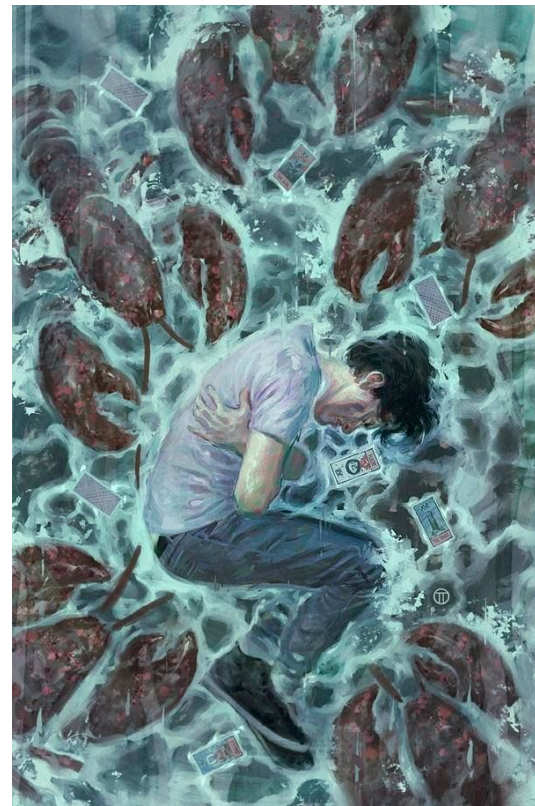


Problems continued

Knowledge-based pronominal coreference

[The city council] refused [the women] a permit because they feared violence.

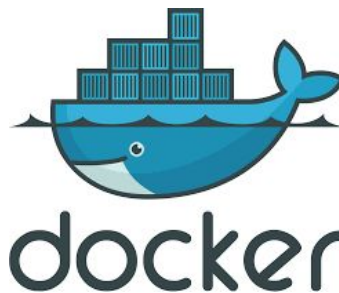
“They” could refer to **the city council** or **the women**. This co-reference can not be determined without additional context.



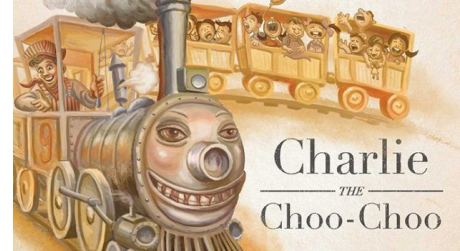
Tech



- 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



CoreNLP



CoreNLP is made up of many annotators. Annotators flag pieces of text with metadata, this metadata will later be built upon to do progressively more complex things.

For this project I used:

- NamedEntityAnnotator (NER) - Finds named entities (NE).
- CorefAnnotator - Finds maps noun phrases (NP) to named entities (NE).
- QuoteAnnotator - Finds quotations, tracks sources.
- CharacterFeatureAnnotator - A custom annotator for collecting information.
- PhraseAnnotator - Annotates animate noun phrases.

Training Data

I used training data built from Project Gutenberg sources and collected by David Bamman, a professor at UC Berkley.

The training data is broken into

- animate unigrams
- female unigrams
- male unigrams
- names associated with genders



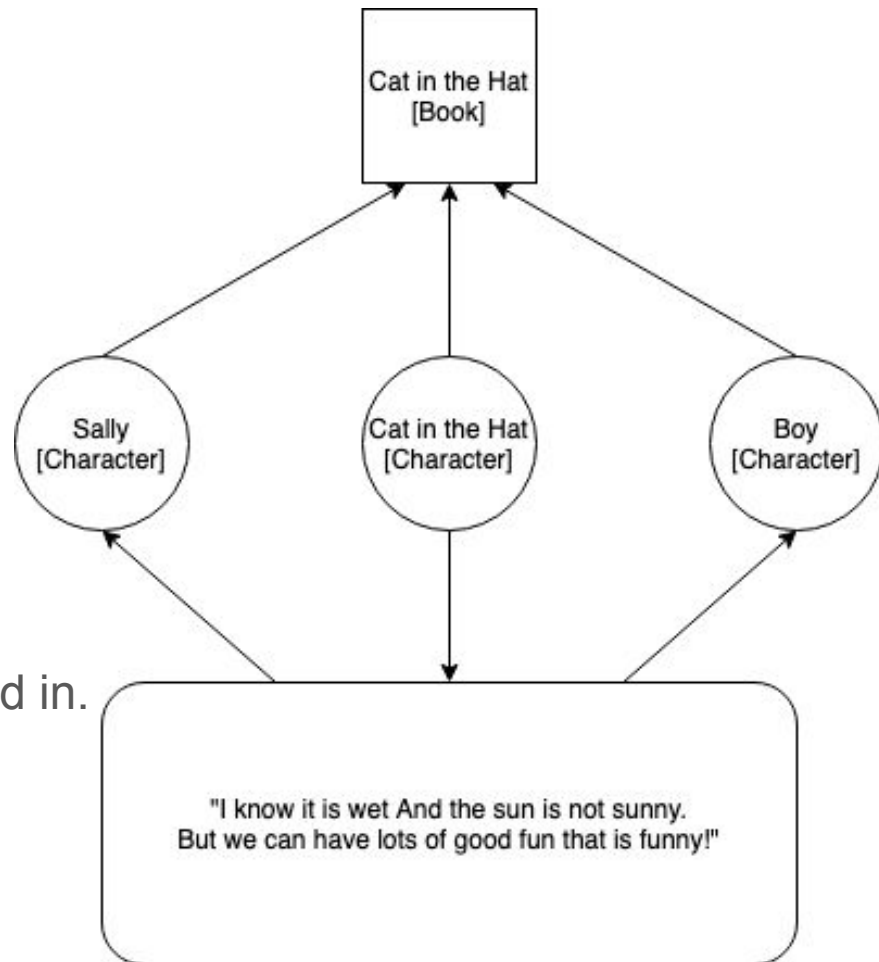
Storage Model

I used Neo4J to collect the data.

Included in the database:

- Characters
- Dialog between multiple characters
- Sentiment of dialog

All nodes link to books they are referenced in.



Cleanup

All together the Dark Tower + Salems Lot is ~5.5k pages. There was a lot of data and most of it was bad.

- Characters are not correctly mapped between books/chapters.
- Characters are referenced few times.
- Nicknames are used/ambiguous character references.



Visualization

I used D3 to make my own graph that used a force layout that pushes until all nodes have a uniform distance.

Each Character Node by default will push all other nodes away. However when you add references to edges the push gets weaker.

Book Nodes also push by default but they can have a weak pull for strong relationships.

All Nodes are added to the graph and then the “physics” is simulated until the graph reaches an equilibrium.



There they stood, ranged along the hill-sides, met
To view the last of me, a living frame
For one more picture! in a sheet of flame
I saw them and I knew them all. And yet
Dauntless the slug-horn to my lips I set,
And blew "*Childe Roland to the Dark Tower came.*"

- Robert Browning

Success!

