Causal DAG Extraction from 3 Short Stories and 3 Movie Scripts

Robert R. Tucci
tucci@ar-tiste.com

April 23, 2023

Abstract

We improve a previously proposed algorithm for doing causal DEFT (DAG Extraction from Text), and then we apply the new algorithm to 2 usecases: 3 short stories by P.G. Wodehouse and 3 movie scripts by Pixar/Disney. The software used to accomplish this endeavor is called "Mappa Mundi" and is available as open source at GitHub.

In this paper, I improve an algorithm for doing causal DEFT (DAG Extraction from Text) that was first proposed in Ref.[3] I then apply the new algorithm to 2 usecases:

1. 3 short stories by P.G. Wodehouse (the text from these was obtained from the Project Gutenberg website [2])

   - Bill the Bloodhound
   - Extricating Young Gussie
   - Wilton's Holiday

2. 3 movie scripts by Pixar/Disney. (the text for these was obtained from the IMSDb website Ref.[1])[1]

   - Toy Story
   - Up
   - WALL-E

The Python script software that was used to accomplish this endeavor is open source and available at Github (see Ref.[4])

References

[1] Internet Movie Script Database (IMSDb). `https://imsdb.com/`.

[2] Project Gutenberg website. `https://www.gutenberg.org`.

[3] Robert R. Tucci. Causal dag extraction from a library of books or videos/movies. `https://arxiv.org/abs/2211.00486`.

[4] Robert R. Tucci. Mappa Mundi at github. `https://github.com/rrtucci/mappa_mundi`.

[1] The Mappa Mundi repo at GitHub contains a Python script called `downloading.py` that uses the BeautifulSoup package to scrape all the 1100+ movie scripts available at the IMSDb website. My original intention was to apply my algorithm to all of those movie scripts. However, due to lack of hardware resources, I had to settle for just 3 movie scripts.