Thomas McIntyre gem5cm@virginia.edu)
DS 5001
08 May 2022

Provenance:
The Fellowship of the Ring
https://www.kaggle.com/datasets/ashishsinhaiitr/lord-of-the-rings-text

A Storm of Swords
https://archive.org/stream3.AStormOfSwords3.%20A%20Storm%20of%20Swords_djvu.txt

A Clash of Kings
https://archive.org/stream/II.AClashOfKings_201803/II.%20A%20Clash%20of%20Kings_djvu.txt

A Game of Thrones
https://archive.org/stream/1AGameOfThrones/1%20A%20Game%20of%20Thrones_djvu.txt

Peter and Wendy / Peter Pan / Jane Eyre / Great Expectations / Oliver Twist / The Iliad / The Odyssey / Wuthering Heights:
Free eBooks | Project Gutenberg

Location:
https://virginia.box.com/s/zetmzoxzpx5osls0l1aep2jnzgk2j97q

Description:
The corpus consists of a grouping of classic texts (Peter and Wendy / Peter Pan / Jane Eyre / Great Expectations / Oliver Twist / The Iliad / The Odyssey / Wuthering Heights) with genres of historic, gothic, children fantasy, etc. combined with more recent fantasy/adventure texts that have turned into blockbuster tv shows and movies (Game of Thrones and Lord of the Rings). The main goal will be to compare Jane Eyre and Game of Thrones, while also including a variety of classic texts.

Format:
The format of the files were all in plain text format. The internal structure was also just plaintext (no XML). The data tables made thereafter were converted into CSV format.