Singapore University of Technology and Design DH 02 137

Introduction

The male dominance in the literary scenes have been prevalent since centuries ago, given the stereotypical gender roles derived from Christian ideology. From 19th century to now, the social hierarchy of women was changed drastically and more and more equality has been seen in the modern world between male and female. By using digital tools, we are curious to identify the time period in which this transition starts to happen, to quantify the qualitative aspects of these gender-related topics and observe trends that are emerged from these data, all through the eyes of literature by female writers.

Methods



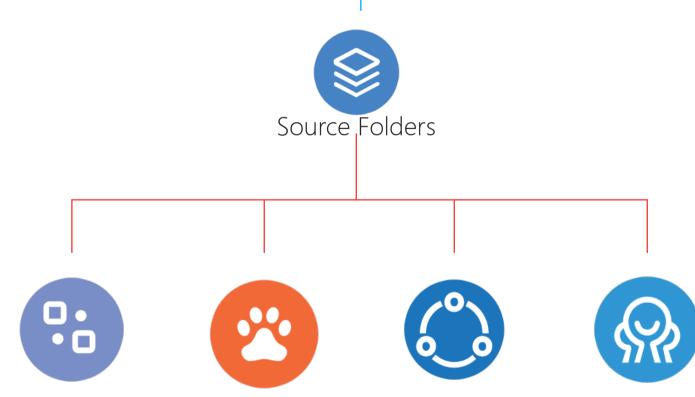








Group text files of books into folders according to different centuries



RStudio Stylo Textblob Mallet Vader Cluster Analysis Sentiment Analysis Topic Modelling Sentiment Analysis

 $-\frac{\partial^2}{\partial t} = -\nabla p + \mu \nabla^2 \vec{v} + \rho \vec{g} + \rho \vec{g$

topics of books

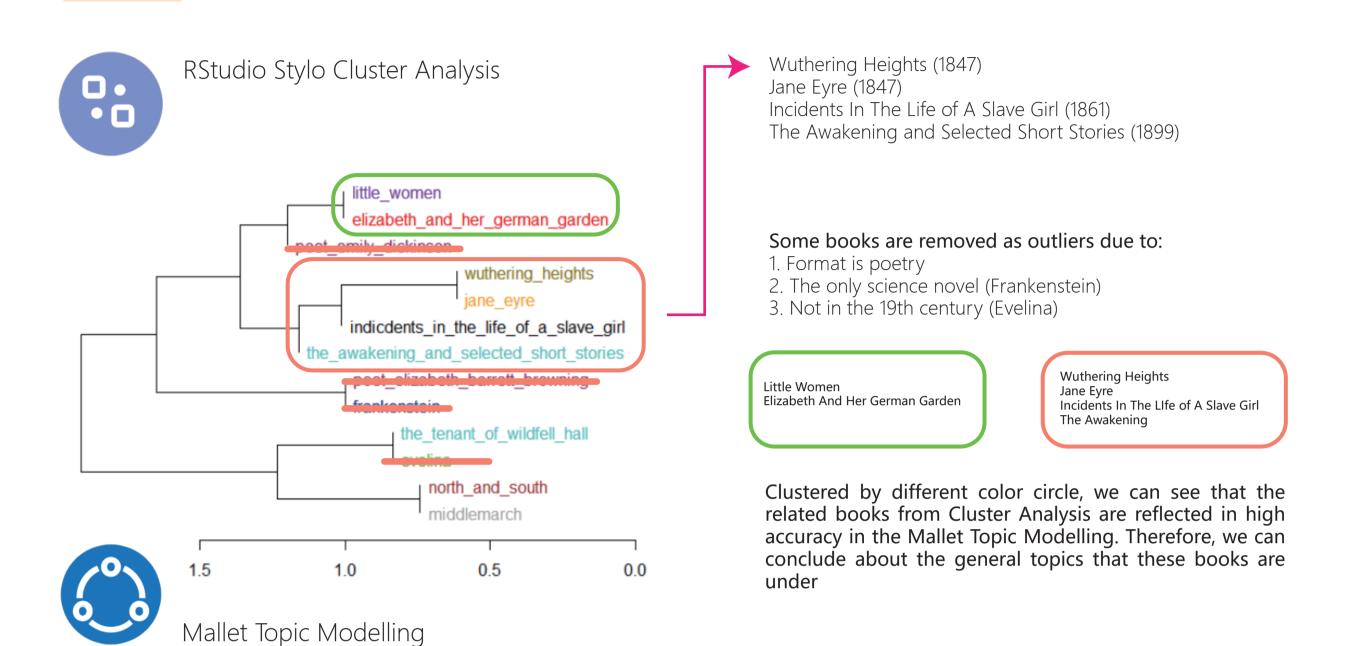
Identify books that have similar characteristics to one another. Preliminary step to draw relations between different books done by different female authors in the same century.

Tokenize text into 10 sentences in each bulk. Plot sentiment polarity score of each bulk into a line graph. Combine all graphs to one result figure

text
Use mallet to conduct unsupervised machine learning to classify books into different topics. Use this together with the cluster analysis data to further categorize the

VadeSentiment to analyse sentiment of short paragraph more accurately. Text is tokenized into 10 sentences in each bulk. Vader outputs "pos" "neu" "neg" and "compound" scores to be displayed in bar graph using matplotlib

Results

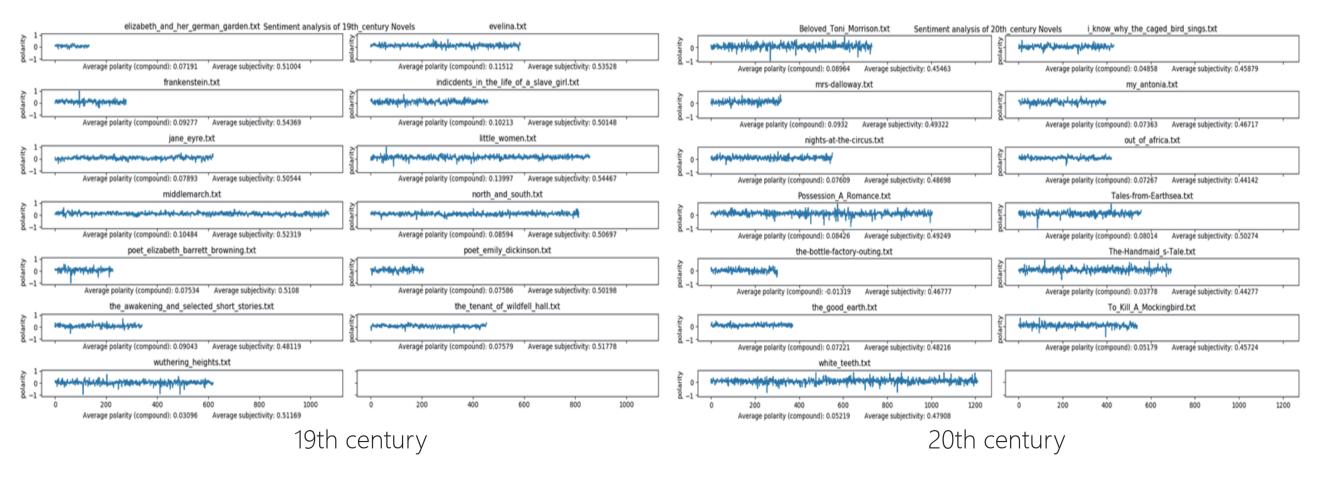


Index	▼ Filename	Undefined	Gender Roles 💌	Gender Inequality	
	0 file:/C:/mallet/mallet/19c/elizabeth_and_her_german_garden.txt	0.07011586	0.206146107	0.723738034	\leftarrow
	1 file:/C:/mallet/mallet/19c/indicdents_in_the_life_of_a_slave_girl.txt	4.5104E-05	0.329462904	0.670491991	\leftarrow
	2 file:/C:/mallet/mallet/19c/jane_eyre.txt	0.067170012	0.387981951	0.544848037	\leftarrow
	3 file:/C:/mallet/mallet/19c/little_women.txt	0.000930974	0.071072863	0.927996162	\leftarrow
	4 file:/C:/mallet/mallet/19c/middlemarch.txt	0.701378465	0.215182181	0.083439354	
	5 file:/C:/mallet/mallet/19c/north_and_south.txt	0.100394232	0.700820958	0.19878481	
	6 file:/C:/mallet/mallet/19c/the_awakening_and_selected_short_stories.txt	0.01194696	0.148150235	0.839902805	\leftarrow
	7 file:/C:/mallet/mallet/19c/the_tenant_of_wildfell_hall.txt	0.04736754	0.751725289	0.200907171	
	8 file:/C:/mallet/mallet/19c/wuthering_heights.txt	0.000477848	0.622121856	0.377400296	\leftarrow



TextBlob Sentiment Analys

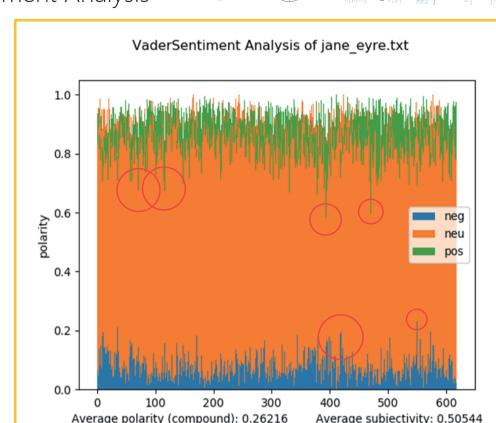


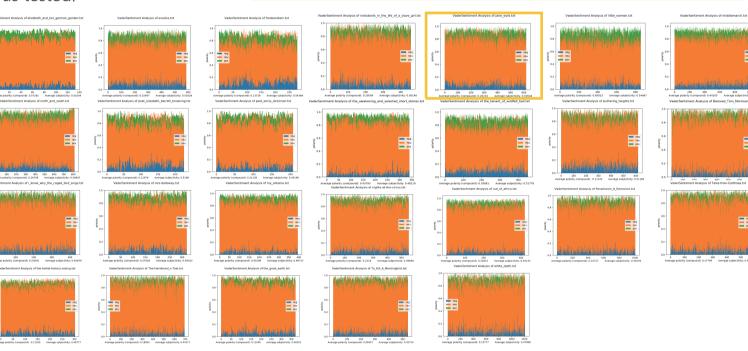


By viewing the average polarity score, it is clear that most of the women writers' novels appear to be neutral or positive. However, since it is the average sentiment of the entire book, it may not be accurate as some parts of the books are just statements and background voices. Therefore, further detailed-breakdown of the sentiment distribution is needed to extract more information.

SP

Valence Aware Dictionary and Sentiment, or VADER for short, was also used to extract the sentiment of each book. Given a string of text, VADER is able to output a value between 0 and 1 as a measure of negativity, positivity and neutrality of a text. As VADER works best with short text, the entire book text is split into sentences using TextBlob. Each sentence was then fed into the sentiment analyzer and plotted as a stacked graph. Bearing in mind that individual sentence may not exhibit sentiments of the author, a bulk size of 10 sentences and 20 sentences





From these detailed-breakdown stacked bar graph of each book, we can see how the sentiment changes throughout the book and identify several points in the book whereby there is a sudden burst of negativity or positivity. We can also tell the flow of sentiments throughout the book by looking at the proportion of the green and blue portions that changes as the content progress.

Sentiment Polarity Score Against Time



The sentiment polarity scores obtained from VaderSentiment Analysis were then plotted against published year of the books. This interesting graph shows a peak near the 1860s and a dip near th 1980s. We can see the up and down of the sentiments expressed out from these novels. Further study could be done to find out what exactly happened during these extreme time that made the sudden change to the way that female writers put in their novels. (i.e. what happened in the 1980s that triggered the author to use aggressive language to get her voice heard?)

Conclusion



From these findings, we identified some major timestamps at which the evolution of women's social status was stimulated, as well as found out the different ways of writing that female authors adopted to get their voices heard and get their messages delivered. However, limitations of few database resources as well as the choice of using average values instead of other more advanced statistical methods, make this project still at its preliminary stage. Further studies are needed to delve into the precise historical events that triggers all these changes.