Book Recommender

Anjali Sebastian

INTRODUCTION

- Book worm
- Analyze & Discover patterns
- Large amount of copyright free literature
- Recommend a similar free book from a excerpt

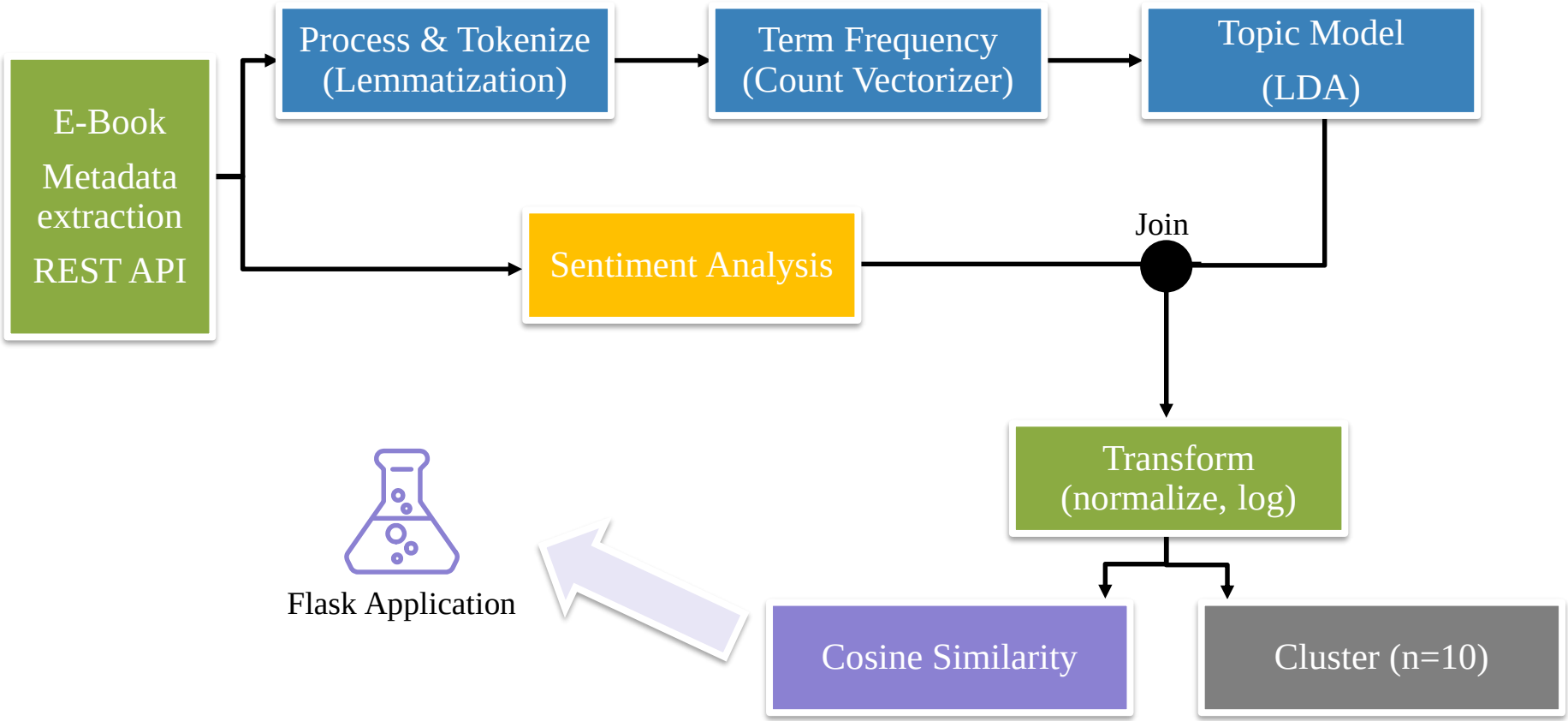
DATA

3000 full e-books from Project Gutenberg - .txt file format

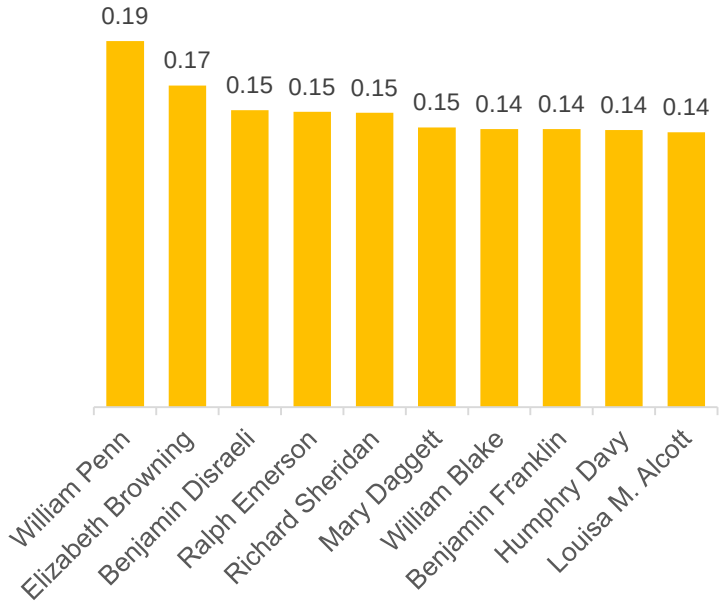
150 authors

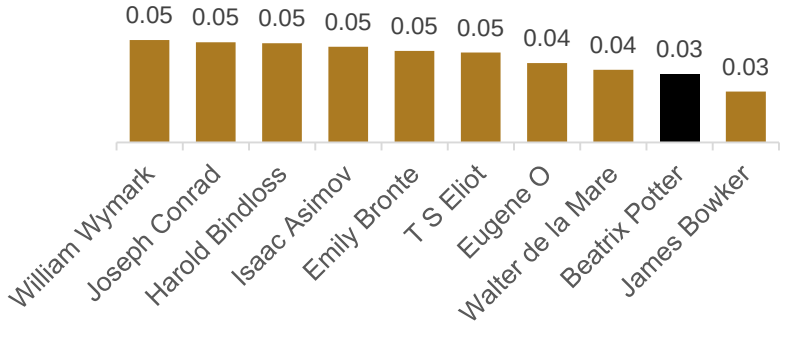
Enhanced dataset of 30,000+ books

DESIGN


E-Book
Metadata extraction
REST API
Process & Tokenize (Lemmatization)
Term Frequency (Count Vectorizer)
Topic Model (LDA)
Sentiment Analysis
Join
Transform (normalize, log)
Cosine Similarity
Cluster (n=10)
Flask Application

SENTIMENT ANALYSIS - POLARITY





PETER RABBIT
THE DARK SIDE

When Benjamin Bunny grew up, **he married his Cousin** Flopsy.

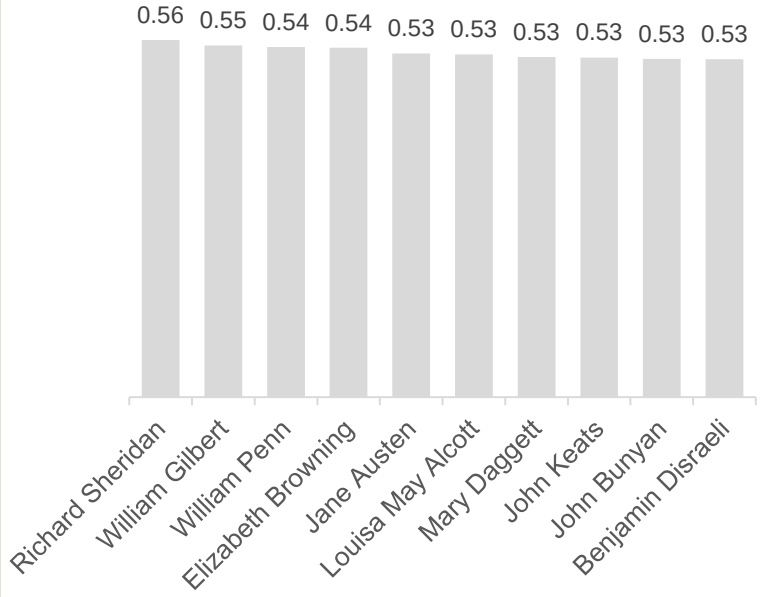
As there was **not** always quite **enough to eat**,—Benjamin used to **borrow** cabbages from Flopsy's brother, Peter Rabbit, who kept a nursery garden.

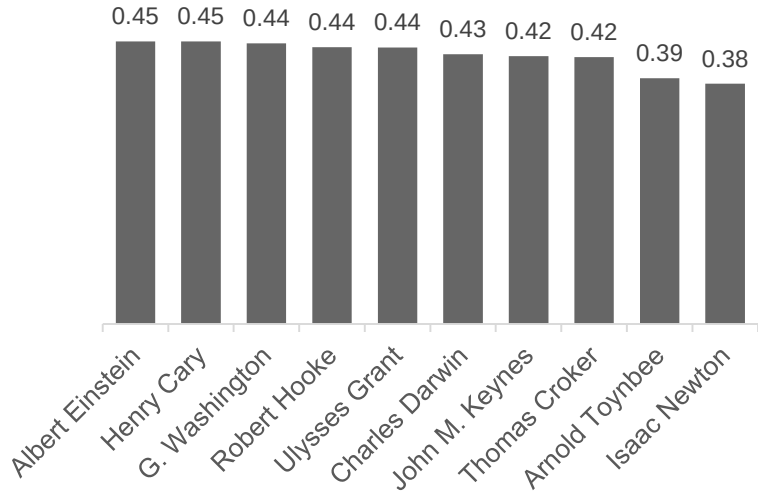
Sometimes Peter Rabbit **had no** cabbages to spare.

SENTIMENT ANALYSIS - SUBJECTIVITY

Abstract

Concrete





POPULAR AUTHORS

Isaac Asimov

Agatha Christie

P G Wodehouse

SMALL SENTENCES !

< 12 words per sentence

TOPIC MODELING

Topic 19 – Religious Text

god , spirit , soul , lord , father , heaven , christ

Sermons, Gospel, Quaker

Charles Kingsley

William Penn

John Bunyan

Topic 10 – Kid's Books

girl , mother , boy , peter , don , cried

The Wonderful Wizard of Oz

The Tale Of Benjamin Bunny

Topic 5 - Poetry

poet , poem , art , nature , line

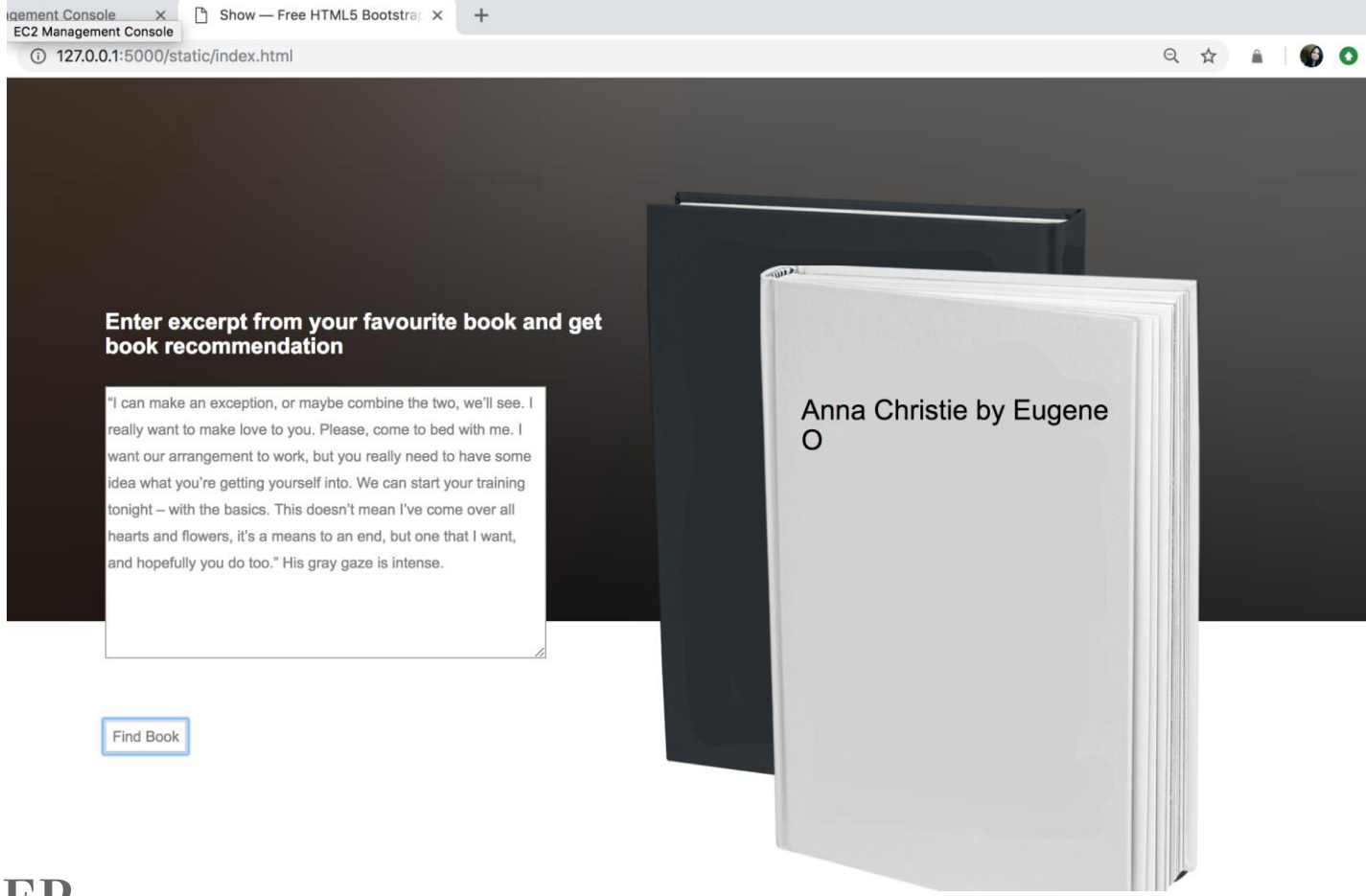
Alexander Pope

Lord Byron

P B Shelley

LDA 20 Topics

Fifty Shades of Grey – Anne Christie



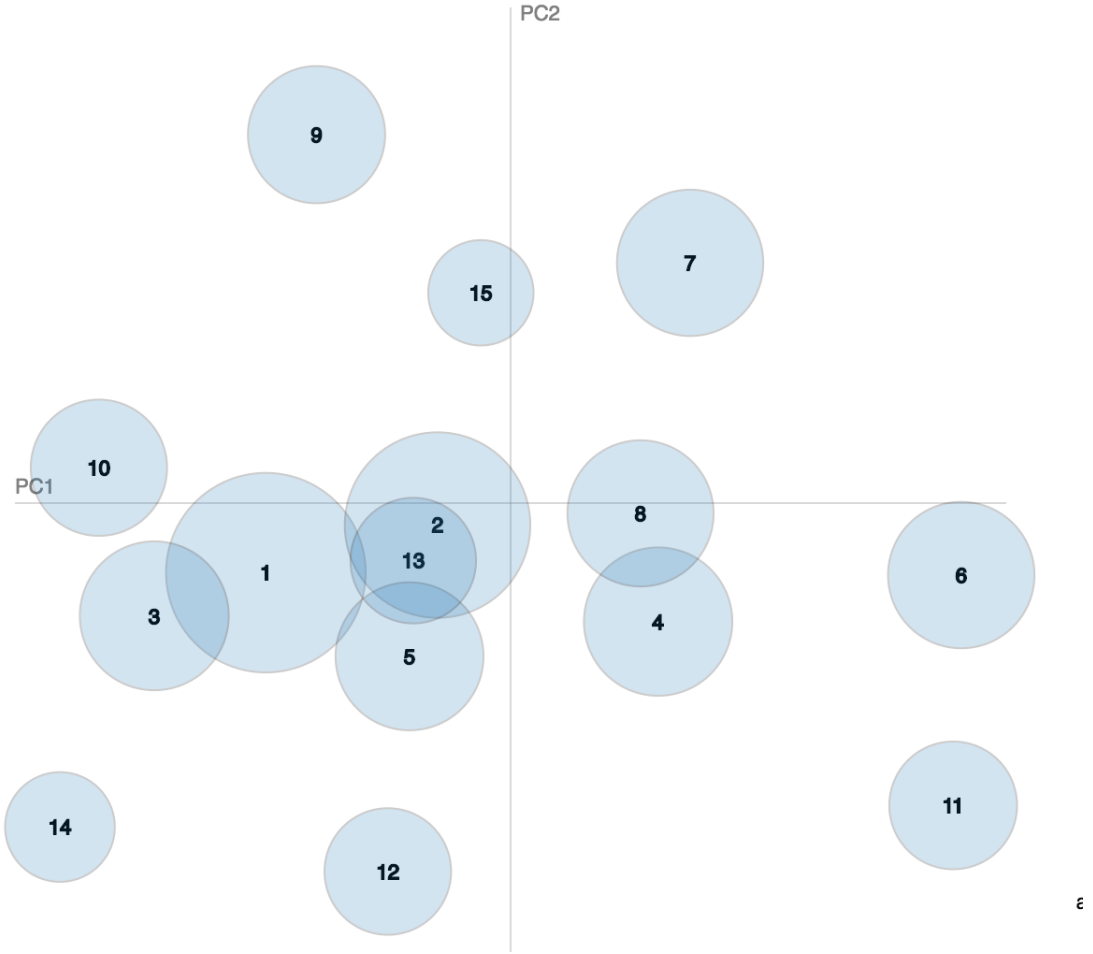
BOOK RECOMMENDER

CONCLUSION & FUTURE WORK

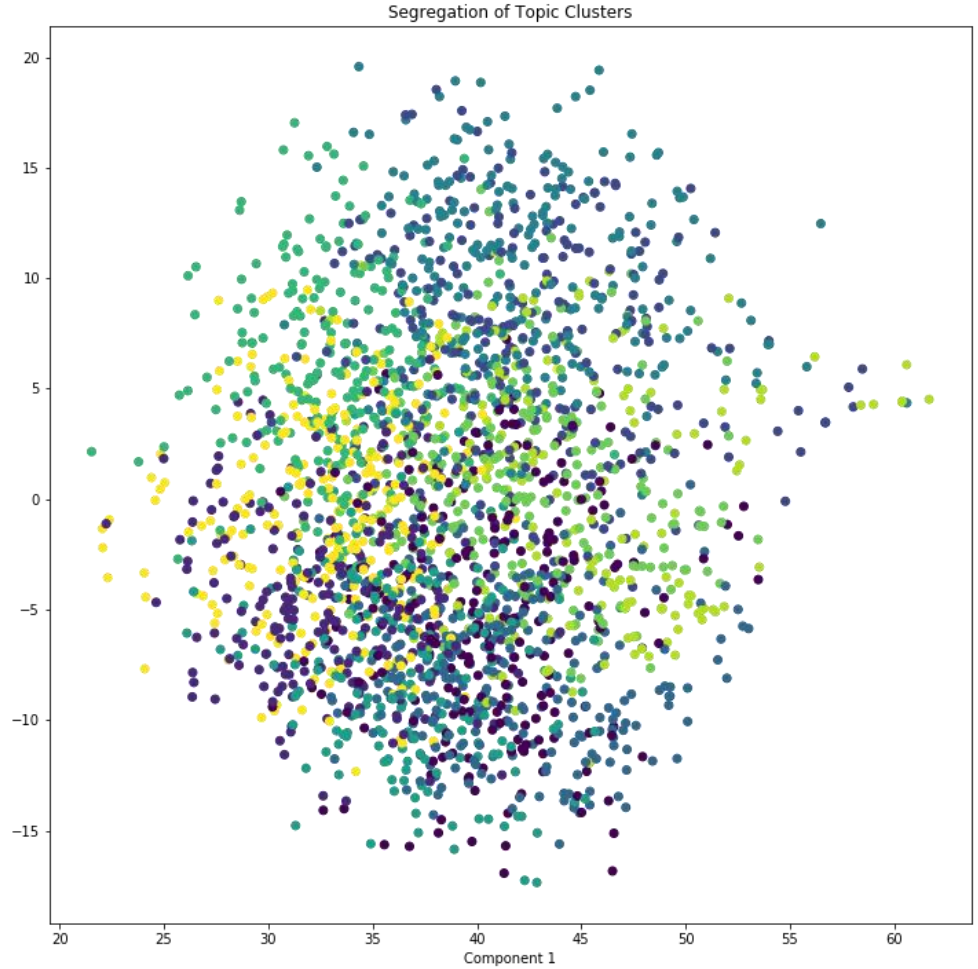
- Able to observe interesting patterns
- Reasonable Suggestions
- Recommendation for fiction vs non fiction
- Bring more context using Neural Networks
- Larger corpus
- Remember user history and adapt

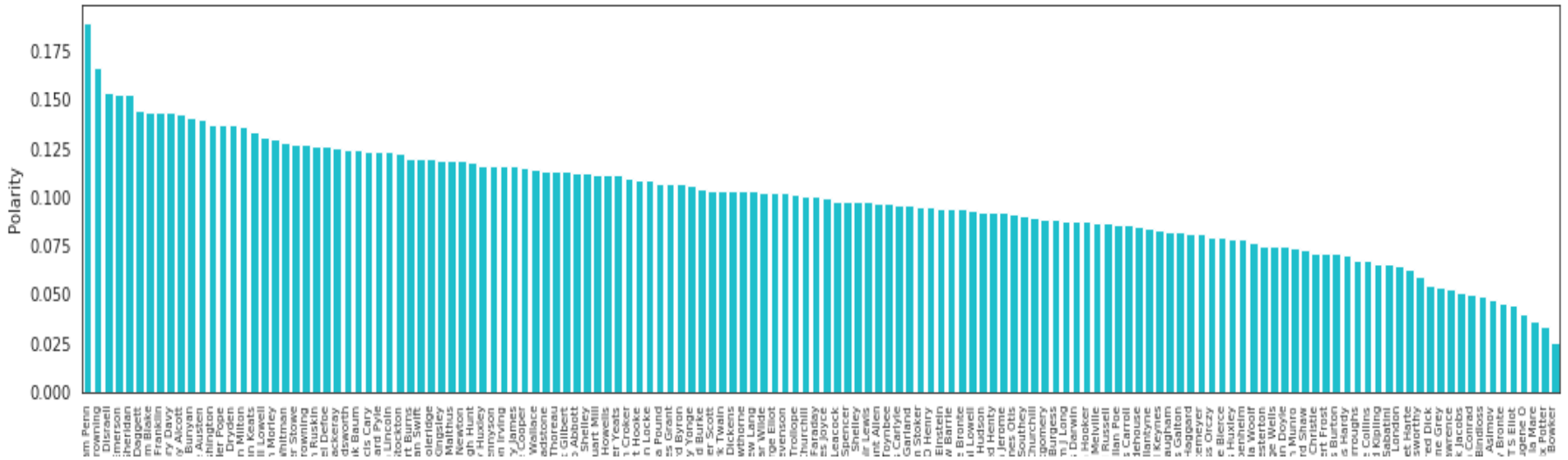
50 - 100 years from now ?

THANK YOU
anjumorris@gmail.com
https://github.com/anjumorris

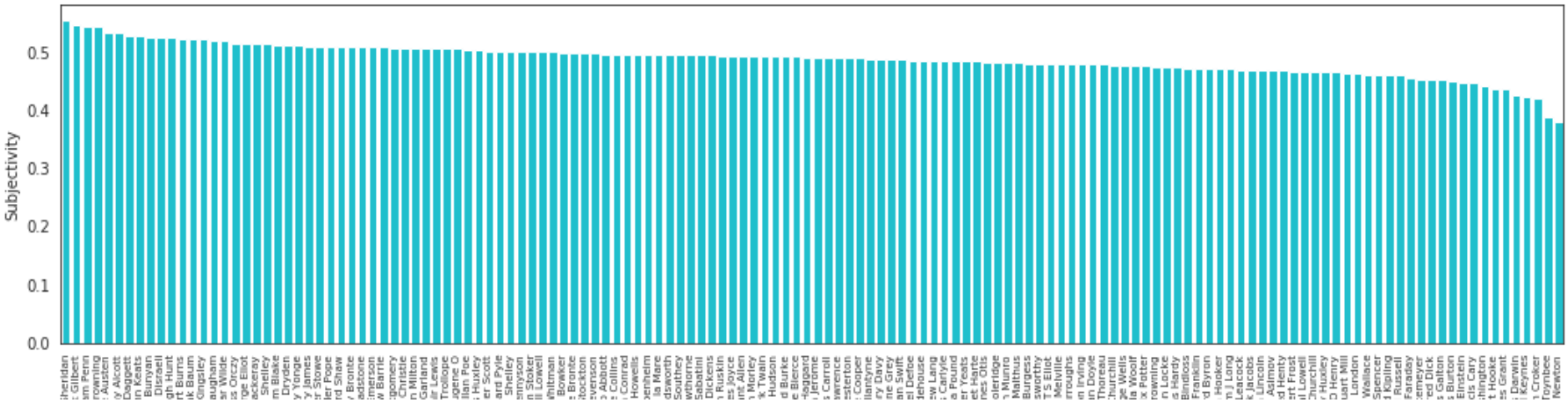

PC2
PC1
9
15
7
10
2
8
1
13
6
3
4
5
14
11
12

CLUSTERING

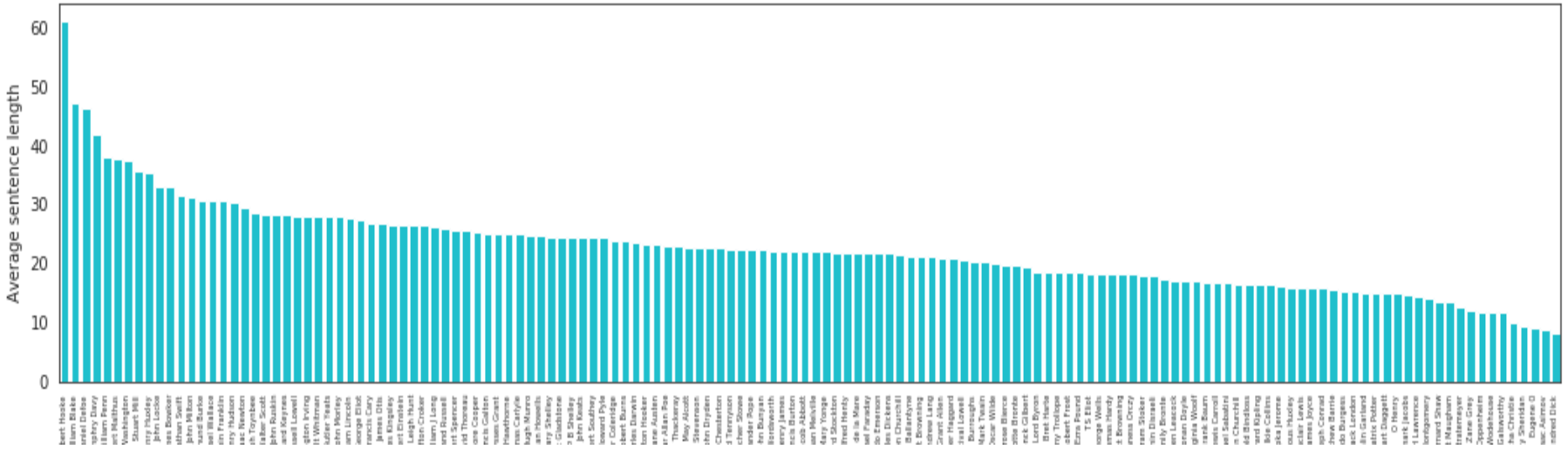

Segregation of Topic Clusters
20
15
10
5
0
−5
−10
−15
20
25
30
35
40
45
50
55
60
Component 1



POLARITY



SUBJECTIVITY



SENTENCE LENGTH