NovelPersepective

Identifying Point of View Characters

Lyndon White, Roberto Togneri, Wei Liu, Mohammed Bennamoun

THE UNIVERSITY OF WESTERN AUSTRALIA

What? Why? Why are you doing this to books?

Many novels, especially epic fantasy, series a written from the Point of View (POV) of many different characters.

They feature multiple parallel stories tracking the journey of each POV character in parallel.

As a reader sometimes one wishes to read just one characters story.

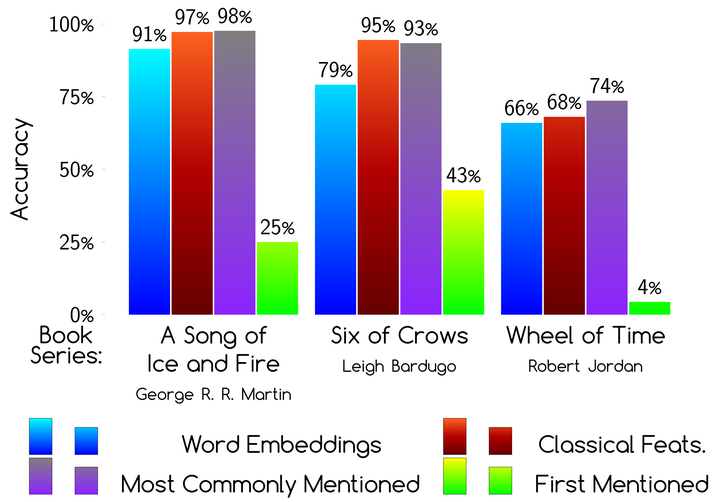
Thus we have made a tool that allows the user to slice-up and restitch their ebooks around a POV character.

The challenging part is that most books to not label the sections with the name of the POV character, rather the reader works it out.

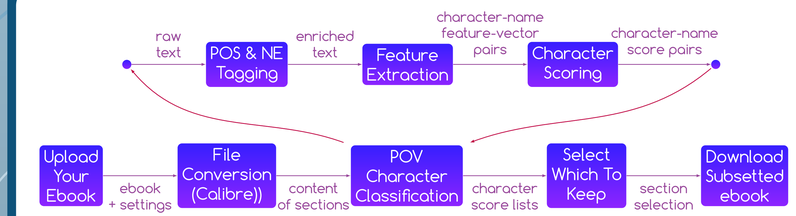
Source Code

Built on CherryPy, NLTK and Scikit-Learn
https://github.com/oxinabox/NovelPerspective
MIT Licensed

Results



Process



Baseline Methods

First Mentioned

Features: First occurrence of name in text
Scoring: Earliest mentioned scores highest, $S_i = 2^{-rank(f_i)}$
Result: Terrible, often other named entities mentioned earlier.

Most Commonly Mentioned

Features: Number of occurrences
Scoring: Most mentioned scores highest, $S_i = \frac{f_i}{\sum_{\forall j} f_j}$
Result: Solid, but fooled by descriptions focusing on others.

Machine Learning Methods

Classical Features

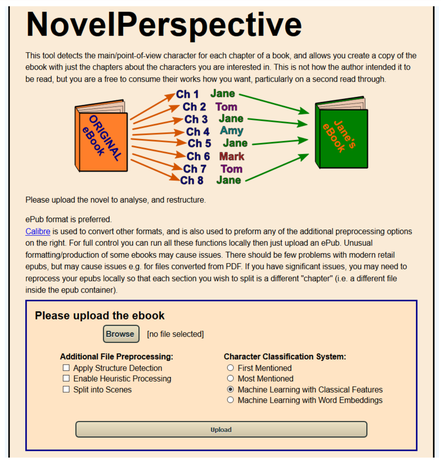
Features: Positional, and occurrence features, plus counts of the part of speech on adjacent words. Total 200 dimensions.
Scoring: use logistic regression model on if POV or not, $S_i = \frac{P(f_i)}{\sum_{\forall j} P(f_j)}$
Result: Really strong. Main characters occur near verbs and grammar.

Word Embeddings

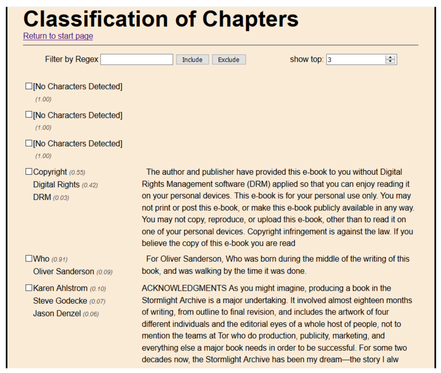
Features: Concatenate the mean of FastText word embeddings for adjacent words Total 600 dimensions.
Scoring: use RBF-SVM model on if POV or not, $S_i = \frac{P(f_i)}{\sum_{\forall j} P(f_j)}$
Result: Really strong, but due to high dimensionality needs sufficient training data.

What does it look like?

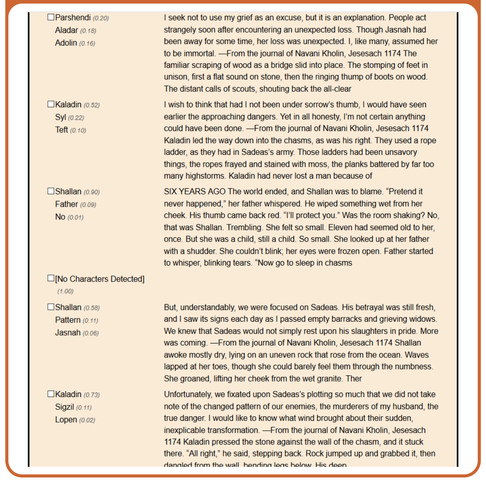
Upload your ebook



Selection Interface



Sections are Labelled by POV



http://novelperspective.ucc.asn.au/

LaTeX TikZposter