

# NovelPersepective

Identifying point of view characters

Lyndon White, Roberto Togneri, Wei Liu, Mohammed Bennamoun

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

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

They feature parallel sub-stories tracking the journey of each POV character.

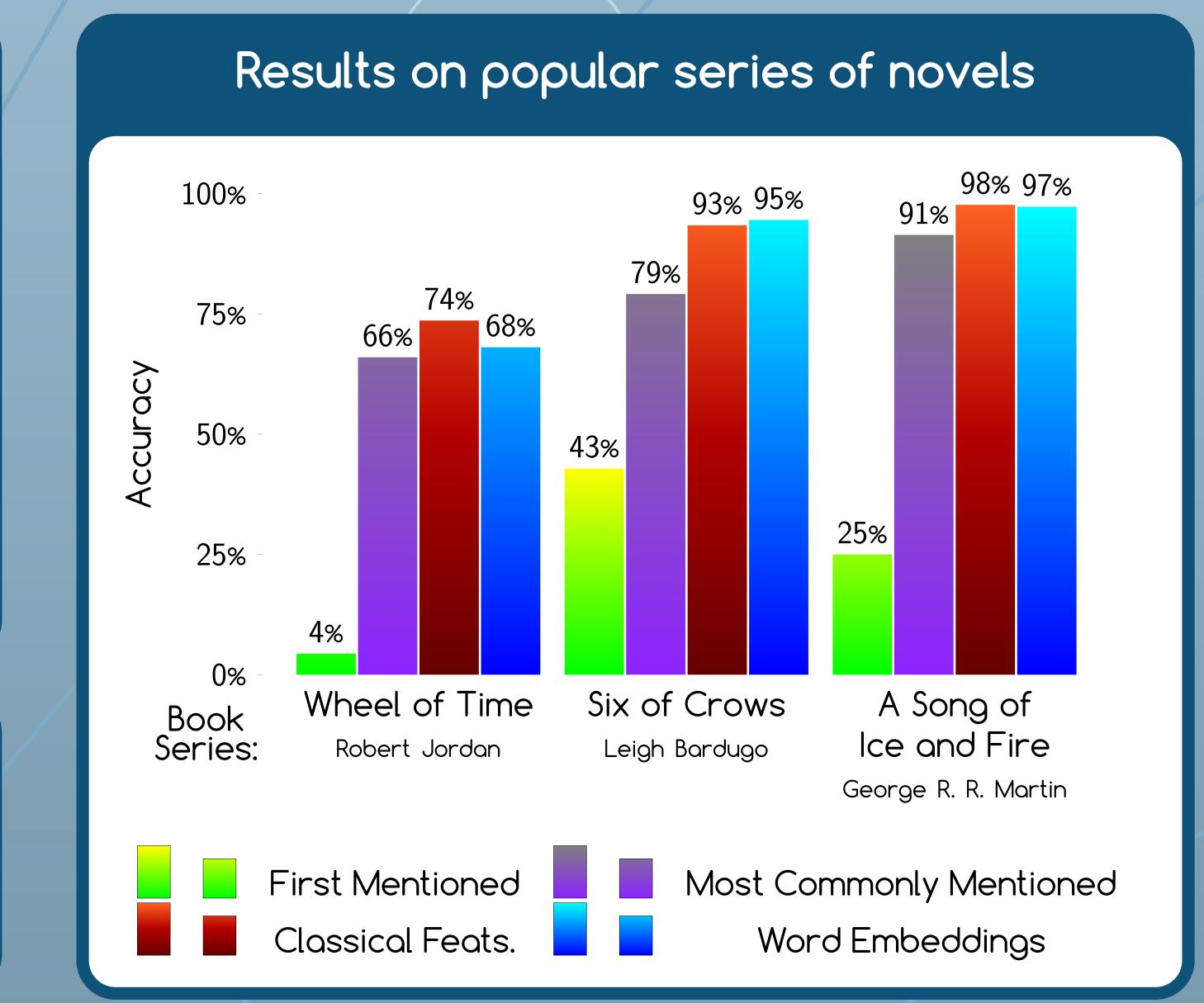
Readers sometimes wish to read just one characters story; particularly for example on a second read through.

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

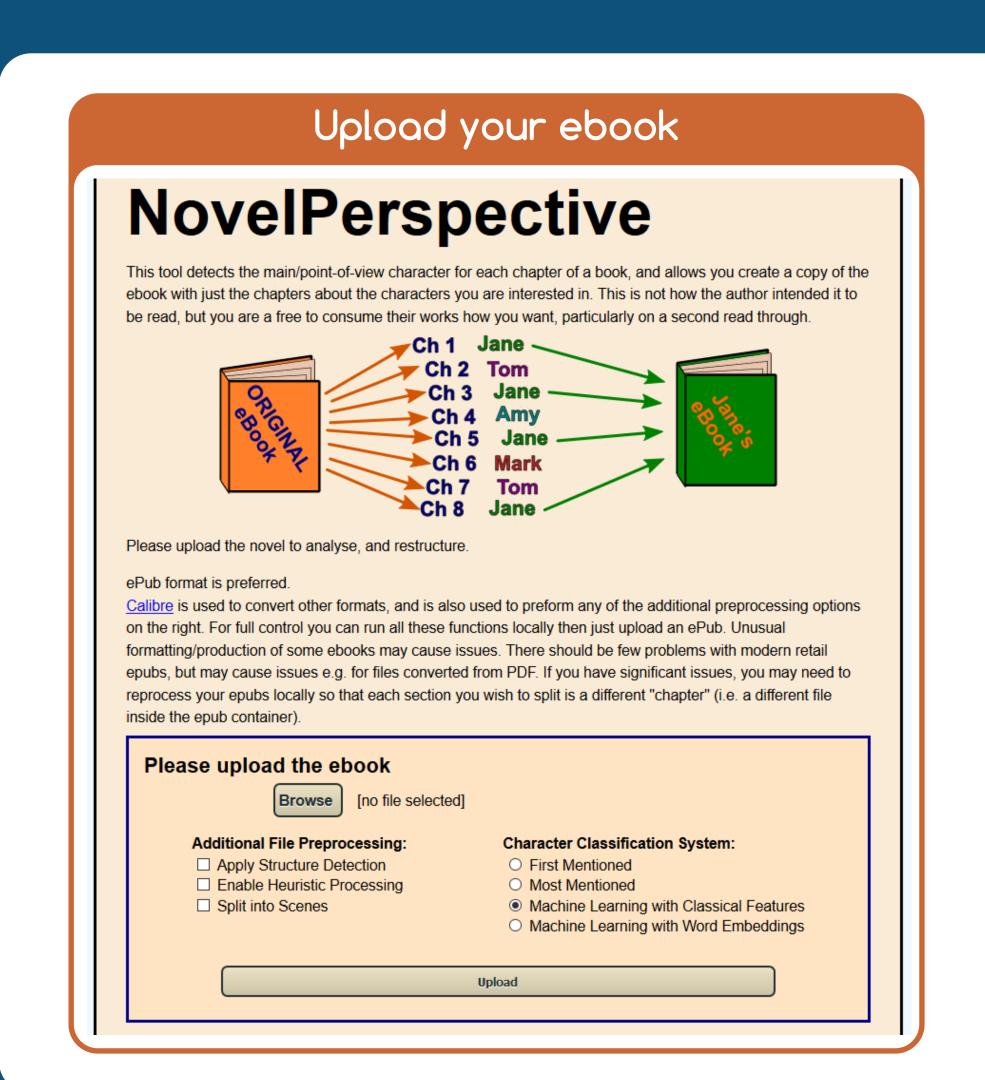
The challenging part is that most books do not label the sections with the name of the POV character, rather the reader works it out.

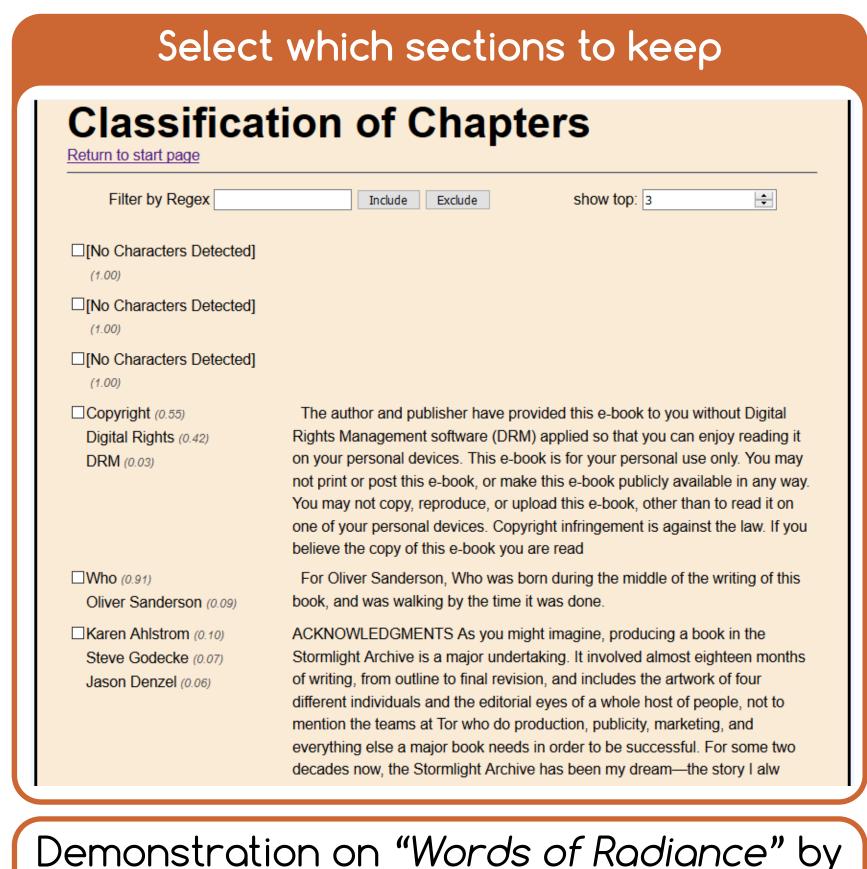
# The source code is publicly available

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



### What does it look like?





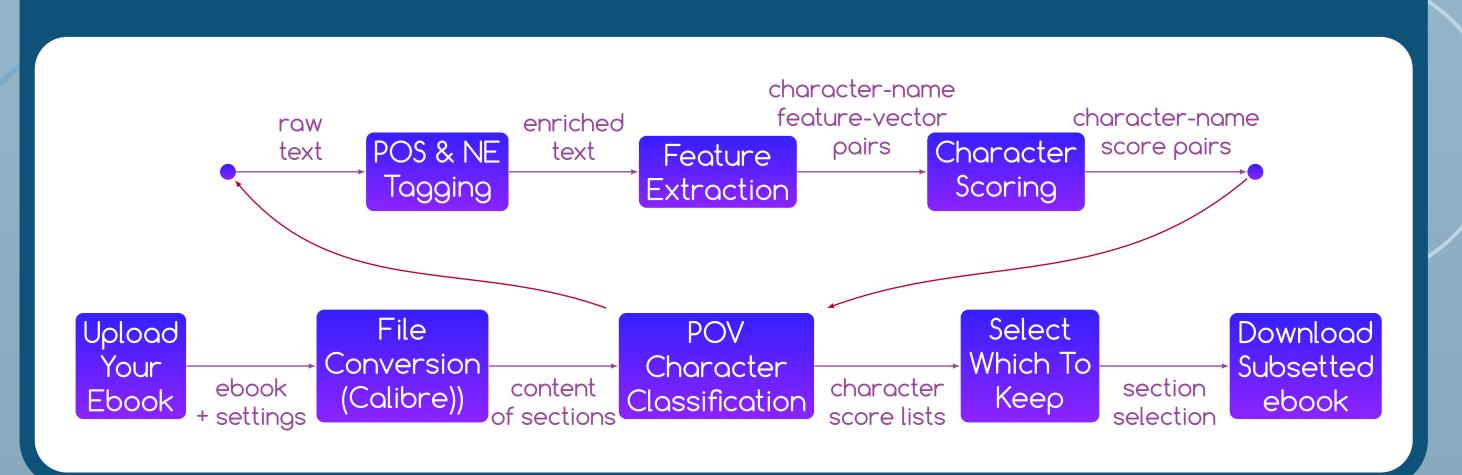
Brandon Sanderson.

Using the classical features method.

seek not to use my grief as an excuse, but it is an explanation. People act Parshendi (0.2 strangely soon after encountering an unexpected loss. Though Jasnah had been away for some time, her loss was unexpected. I, like many, assumed her o be immortal. —From the journal of Navani Kholin, Jesesach 1174 The familiar scraping of wood as a bridge slid into place. The stomping of feet in unison, first a flat sound on stone, then the ringing thump of boots on wood. that was Shallan, Trembling, She felt so small, Eleven had seemed old to her. once. But she was a child, still a child. So small. She looked up at her father with a shudder. She couldn't blink; her eyes were frozen open. Father started to whisper, blinking tears. "Now go to sleep in chasms [No Characters Detected] But, understandably, we were focused on Sadeas. His betrayal was still fresh, and I saw its signs each day as I passed empty barracks and grieving widows. We knew that Sadeas would not simply rest upon his slaughters in pride. More was coming. —From the journal of Navani Kholin, Jesesach 1174 Shallan awoke mostly dry, lying on an uneven rock that rose from the ocean. Waves apped at her toes, though she could barely feel them through the numbness. She groaned, lifting her cheek from the wet granite. Ther Unfortunately, we fixated upon Sadeas's plotting so much that we did not take □Kaladin (0.73) note of the changed pattern of our enemies, the murderers of my husband, the true danger. I would like to know what wind brought about their sudden, nexplicable transformation. —From the journal of Navani Kholin, Jesesach 1174 Kaladin pressed the stone against the wall of the chasm, and it stuck there. "All right," he said, stepping back. Rock jumped up and grabbed it, then dangled from the wall, bending legs below. His deen

Sections are labelled by POV & score

# The process for subsetting ebooks by POV



# Baseline methods for determining POV

#### First Mentioned Named Entity

**Features**: first occurrence of named entity token in the section. **Scoring**: earliest mentioned scores highest,  $S_i = 2^{-rank(f_i)}$  **Result**: **terrible**. Other named entities often occur before POV.

#### Most Commonly Mentioned Named Entity

**Features**: count of occurrences of named entity token in section. **Scoring**: most mentioned scores highest,  $S_i = \frac{f_i}{\sum_{j \in f_i}}$ 

Result: generally solid, but fooled by descriptions focusing on others.

# Machine learning methods for determining POV

#### Classical Features + Logistic Regression

**Features**: position, and occurrence frequency, plus parts of speech cooccurring frequency. Total 200 dims. **Scoring**: use logistic regression model on if POV or not,  $S_i = \frac{P(f_i)}{\sum_{\forall j} P(f_j)}$ 

**Result**: generally great. Main characters occur near verbs and grammar. This gives an edge over frequency information alone.

#### Word Embeddings + RBF-SVM

**Features**: concatentation of FastText word embeddings for the adjacent words, averaged over all occurrences. Total 600 dims. **Scoring**: use RBF-SVM model on if POV or not,  $S_i = \frac{P(f_i)}{\sum_{M} P(f_i)}$ 

Result: generally great. However, due to high dimensionality needs sufficient training data from other labelled books.