Fletcher Investigation Team

• • •

Sept 21, 2020

Team

- Jerry Duncan - Project Manager / Data Engineer

- Everett Rush - Data Scientist

- Daniel Schultz - Presentation Slides

- Quan Zhou - Data Scientist

- All - Reporting

Fletcher Books

- The Middle of Things - 272 Pages

- Scarhaven Keep - 216 Pages

- The Talleyrand Maxim -182 Pages

- The Middle Temple Murder - 256 Pages

- Dead Men's Money - 289 Pages

Methods

- Import each novel from URL (Beautiful Soap)

- Split sentences for each book using regular expressions (RE)

    - NLTK for tokenization: Chapters -> Sentences

- Custom REGEX formula for processing

    - Extract data into JSON format

- Graph Relevant results using PyPlot

Results

Goal: Answer some key questions.

1. When does the detective (or a pair) occur for the first time?

2. When is the crime first mentioned - the type of the crime and the details?

3. When is the perpetrator first mentioned?

4. What are the 3 words that occur around the perpetrator on each mention?

5. When and how the detective/detectives and the perpetrators co-occur?

6. When are other suspects first introduced?

The Middle of Things

1. When does the detective (or a pair) occur for the first time?
   a. Detective Viner: Chapter 1 Sentence 1
2. When is the crime first mentioned - the type of the crime and the details?
   a. Murder: Chapter 1 Sentence 76
3. When is the perpetrator first mentioned?
   a. Cortelyon: Chapter 2 Sentence 54
4. What are the 3 words that occur around the perpetrator on each mention?
   a. 90 context words
5. When and how the detective/detectives and the perpetrators co-occur?
   a. Detective-perp Co-occurrences [(3, 78), (28, 134)]
6. When are other suspects first introduced?
   a. Penkridge: Chapter 1 Sentence 1

Scarhaven Keep

1. When does the detective (or a pair) occur for the first time?
    a. Detective Copplestone: Chapter 1 Sentence 39
2. When is the crime first mentioned - the type of the crime and the details?
    a. Fraud: Chapter 15 sentence 139
3. When is the perpetrator first mentioned?
    a. Peter Chatfield, Addie Chatfield: Chapter 4 Sentence 114
4. What are the 3 words that occur around the perpetrator on each mention?
    a. 969 context words
5. When and how the detective/detectives and the perpetrators co-occur?
    a. Detective-perp Co-occurrences [(6, 3), (6, 32), (8, 161), (18, 73), (21, 47), (21, 84), (21, 98), (22, 87), (23, 45), (27, 37), (29, 110), (30, 189), (31, 1), (31, 78), (31, 160)]
6. When are other suspects first introduced?
    a. Audrey Greyle: Chapter 3 Sentence 166

The Talleyrand Maxim

1. When does the detective (or a pair) occur for the first time?
   a. Detective Linford Pratt: Chapter 1 Sentence 1
2. When is the crime first mentioned - the type of the crime and the details?
   a. Murder: Chapter 7 Sentence 93
3. When is the perpetrator first mentioned?
   a. Harper Mallathorpe: Chapter 1 Sentence 57
4. What are the 3 words that occur around the perpetrator on each mention?
   a. 635 context words
5. When and how the detective/detectives and the perpetrators co-occur?
   a. Detective-perp Co-occurrences [(2, 130), (2, 147), (2, 148), (2, 149), (2, 150), (5, 1), (5, 18), (5, 21), (5, 26), (5, 27), (5, 30), (5, 114), (5, 117), (5, 158), (5, 171), (5, 193), (6, 121), (7, 122), (8, 1), (8, 11), (8, 21), (8, 24), (8, 35), (8, 50), (8, 97), (8, 165), (10, 90), (10, 164), (10, 182), (11, 40), (11, 74), (12, 168), (12, 179), (12, 217), (12, 219), (12, 253), (12, 270), (12, 278), (12, 291), (12, 299), (12, 300), (12, 315), (13, 32), (13, 34), (13, 68), (13, 85), (13, 93), (13, 106), (13, 114), (13, 115), (13, 130), (14, 46), (14, 154), (15, 129), (16, 113), (16, 138), (16, 156), (16, 173), (16, 178), (17, 51), (17, 56), (17, 67), (18, 148), (18, 150), (21, 10), (21, 13), (21, 70), (21, 125), (21, 154), (22, 1), (25, 34), (25, 40), (27, 48), (27, 50), (27, 63)]
6. When are other suspects first introduced?
   a. Nesta: first Chapter 2 Sentence 17

The Middle Temple Murder

1. When does the detective (or a pair) occur for the first time?
   a. Detective Frank Spargo: Chapter 1 Sentence 1
2. When is the crime first mentioned - the type of the crime and the details?
   a. Murder: Chapter 1 Sentence 24
3. When is the perpetrator first mentioned?
   a. Jane Baylis: Chapter 19 Sentence 24
4. What are the 3 words that occur around the perpetrator on each mention?
   a. 360 context words
5. When and how the detective/detectives and the perpetrators co-occur?
   a. Detective-perp Co-occurrences [(19, 41), (21, 153), (23, 11), (23, 13), (23, 34), (23, 35), (24, 1), (24, 22), (25, 17), (26, 31), (27, 6), (27, 8), (27, 17), (27, 59), (29, 71), (30, 133), (36, 131)]
6. When are other suspects first introduced?
   a. Chamberlayne: Chapter 6 Sentence 156

Dead Men's Money

1. When does the detective (or a pair) occur for the first time?
   a. Detective Hugh Moneylaw: Chapter 1 Sentence 19
2. When is the crime first mentioned - the type of the crime and the details?
   a. Murder: Chapter 1 Sentence 3
3. When is the perpetrator first mentioned?
   a. Gilbert Carstairs: Chapter 7 Sentence 6
4. What are the 3 words that occur around the perpetrator on each mention?
   a. 676 context words
5. When and how the detective/detectives and the perpetrators co-occur?
   a. Detective-perp Co-occurrences [(22, 67)]
6. When are other suspects first introduced?
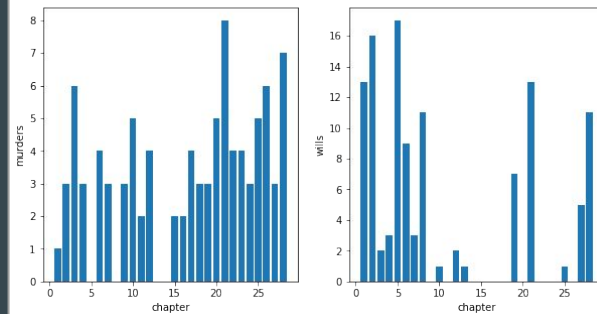   a. Dead Men's Money had no Other Suspects

Graphs
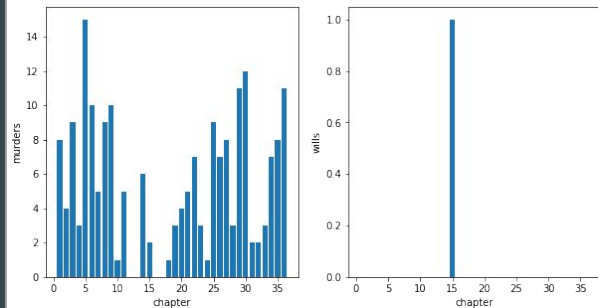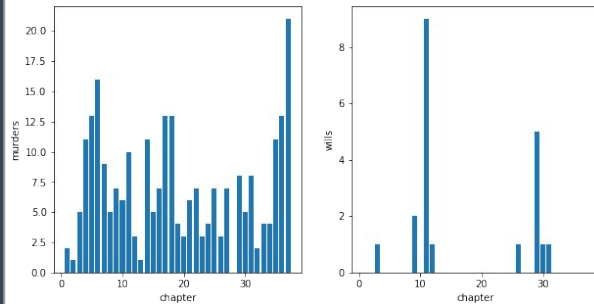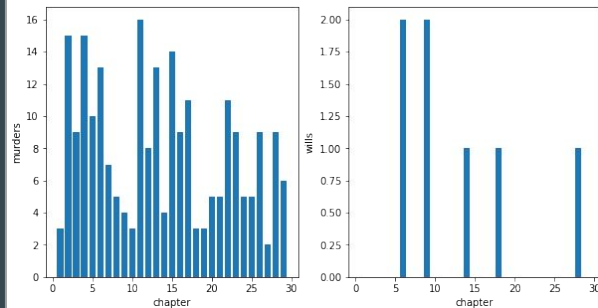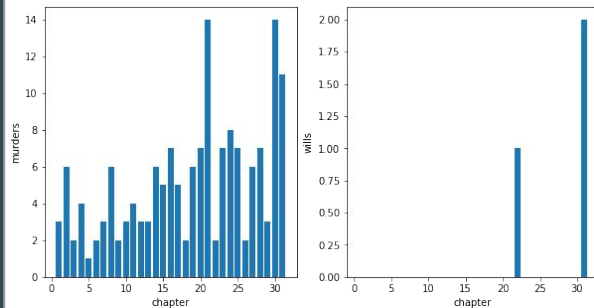
RegEx

```python
@property
def sentences(self):
    '''
    Splits the chapter text into sentences. This is a hard task so some decisions need to be made.

    Case 1: (?<!\w\.\w.)(?<![A-Z][a-z]\.)(?<![A-Z][a-z][a-z]\.)(?<=[.?!])\s
        (?<!\w\.\w.): Negative lookbehind to prevent splitting on i.e. and e.g.
    Python requires fixed width patterns for lookbehinds, so we have split these
        (?<![A-Z][a-z]\.): Negative lookbehind to prevent splitting on Mr.
        (?<![A-Z][a-z][a-z]\.): Negative lookbehind to prevent splitting on Mrs.
        (?<=[.?!]): Positive lookbehing to make sure we're only splitting after ., ?, or !
        \s: Any white-space character
    Case 2: (?<=[.?!][\"])\s(?=[\"A-Z])
        (?<=[.?!][\"]): Positive lookbehind to make sure we're only splitting after .", ?", or !"
        \s: Any white-space character
        (?=[\"A-Z]): Positive lookahead to make sure we're only splitting before " or a capital letter

    Examples:
        ...a fashion which had become a habit. Miss Penkridge...
            Sentence 1: ...a fashion which had become a habit.
            Sentence 2: Miss Penkridge...
        ...content. "So he did it! Now, I should never have thought it! The last person...
            Sentence 1: ...content.
            Sentence 2: "So he did it!
            Sentence 3: Now, I should never have thought it!
            Sentence 4: The last person...
        ...this sort of stuff?" "Stuff?" demanded Miss Penkridge, who had resumed her...
            Sentence 1: ...this sort of stuff?"
            Sentence 2: "Stuff?" demanded Miss Penkridge, who had resumed her...
    '''
    splits = re.split('(?<!\w\.\w.)(?<![A-Z][a-z]\.)(?<![A-Z][a-z][a-z]\.)(?<=[.?!;])\s|(?<=[.?!][\"])\s(?=[\"A-Z])', self.text)
    return zip(range(1, len(splits)+1), splits)
```
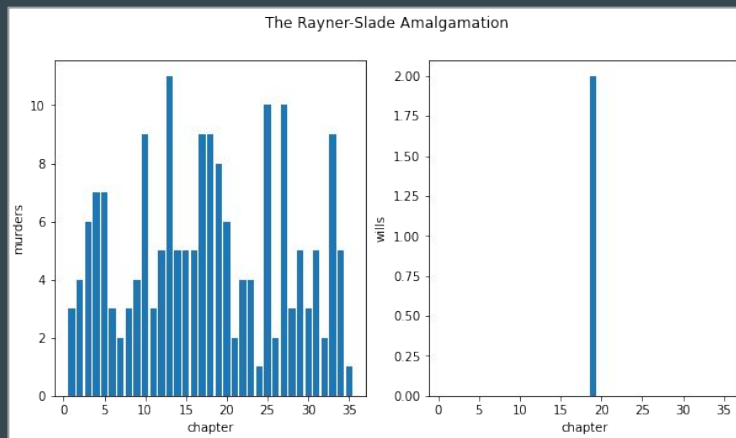
What Worked/Didn't Work

- Plot complexity can be an issue.
    - The Rayner-Slade Amalgamation - Very complex plot & multiple personalities

Questions

Interactive Demo - https://ece692-nlp.web.app/