ANALYSIS OF FIVE OF J.S. FLETCHER'S MURDER MYSTERY NOVELS

Jerry Duncan, Everett Rush, Daniel Schultz, and Quan Zhou
Electrical Engineering and Computer Science (EECS)
The University of Tennessee
Knoxville, TN 37996
jdunca51@vols.utk.edu, erush3@vols.utk.edu, dschult9@vols.utk.edu, qzhou10@vols.utk.edu

September 23, 2020

ABSTRACT

Natural Language Processing is one of the largest and fastest growing Machine Learning fields. Its applicability to such a large variety of problems, makes it both versatile and often used to solve any problem relating to text data. In this work we use very simplistic Natural Language Processing to analyze the plot structure of five different murder mystery books written by J.S. Fletcher. We use Beautiful Soup to scrape the books off of Project Gutenberg, split the books into chapters and sentences using Regular Expressions, and then search through the corpora for answers to six plot-defining questions. From these we create a good view of how each novel's plot structure relates to each other.

***K*eywords** Natural Language Processing · Plot Structure Analysis · Regular Expressions

1 Introduction

Machine learning techniques have been applied in many scientific domains such as healthcare, biology, chemistry, phsyics, etc. Other fields such as humanities have a unique set of challenges and a much fewer applications. One such challenge is that everyone has their own opinion about the quality and interpretation of a work. Some people may enjoy and highly praise one work while others may not. Furthermore, because art is made for humans, art usually invokes a human reaction or emotion. For example, an author may add an element of suspense into a work of fiction or have a comic moment. These qualities can be difficult to describe to other humans, let alone a computer.

Plot structure analysis is a task that does not immediately translate well to a machine learning problem. Plot structure analysis plots the main sequences of events in a work of fiction on a timeline. Often the context surrounding these main events is described as well. Writers and critics usually perform this task by hand as they analyze a work of fiction. Plots often contain surprises and twists that catch readers off-guard. Characters often change and evolve throughout the course of a novel as well. For example a friendly character could turn into a criminal or the least suspecting person could have committed the gravest of crimes.

In this work we perform a computational analysis of the plot structure of five J.S. Fletcher novels. Each novel contains a detective that goes about solving a crime or supposed crime. Many times, the detective is not a detective by trade but gets dragged into an investigation due to circumstances. The novels always end by tying up all the loose ends and solving the crime. Often times, some unexpected events happen along the way such as the suicide of the lead character.

2 Approach

Data Sources We selected five murder mysteries by J.S. Fletcher for analysis: Dead Men's Money [1], The Middle of Things [2], The Middle Temple Murder [3], The Talleyrand Maxim [4], and Scarhaven Keep [5]. His works are freely available through Project Gutenberg [6]. We selected these five murder mysteries because of their conformity to standard literary tropes and their potential to reveal hidden structure in Fletcher's works. In each book, there is always an investigator, a murderer(s), and a timeline from the discovery of a murder to the identification of the murderer(s). However, this does not imply that they all have the same, overly rigid structure, as we found that some of the novels contain complex subplots. For instance, a killer might have multiple pseudonyms, receive assistance from a number of co-conspirators, or be revealed to not be a murderer at all. Very often the murder is enacted to support a larger objective such as the concealment of a second will that removes the killer from receiving their inheritance.

Data Preparation For The Tallyrand Maxim specifically, we used the Stanford Core NLP Part-of-Speech (POS) tagger from the Natural Language TookKit (NLTK) to help identify instances where the term *will* is used as a noun rather than a verb, to prevent false positives that do not refer to the murderer's goal. We also used a Wordnet synonym set to identify synonyms of *murder* and *will*, all of which were manually transformed into their respective lemmas.

To help split the chapters into sentences that we could further analyze, we used a complex, handcrafted regex expression to split sentences that is better described in Appendix A. We also used NLTK to both identify proper nouns and to split contractions with its Treebank Word Tokenizer.

Background Research We conducted extensive background research to identify the investigator, criminals, and other potential accomplices. This was a manual process because there are no in-depth plot summaries or analyses for any of the J.S. Fletcher works. GoodReads [7] contains the most basic summaries of many books. Kadaxis [8] has a good list of characters. We had to read large portions of the books because our resources were lacking a list of the criminals, the ending, and the other suspects.

Analysis We started by performing regex searches for the first occurrence of when the principle characters appeared in their respective novels, as well as every co-occurrence of the investigator and murderer(s). We also recorded the number of times that *murder* and *will* occurred in The Tallyrand Maxim. And, as noted in **Data Preparation**, we filtered *will* by part of speech to remove instances where it was used as a verb.

3 Results

After extensive investigation of deep reading and manual regex search, the detailed information of the above five Fletcher's novels is shown from Table 1 to Table 5.

Furthermore, using the Regex approach to search corresponding detective, perpetrator and suspects, the results are shown in table 6. And question 1 to question 6 are denoted as 1-6. Below illustrates question 1-6 that we answered for each book.

1. When does the detective (or a pair) occur for the first time - chapter #, the sentence(s) # in a chapter;
2. When is the crime first mentioned - the type of the crime and the details - chapter #, the sentence(s) # in a chapter;
3. When is the perpetrator first mentioned - chapter #, the sentence(s) # in a chapter;
4. What are the 3 words that occur around the perpetrator on each mention (i.e., the three words preceding, and the three words following the mention of a perpetrator);
5. When and how the detective/detectives and the perpetrators co-occur - chapter #, the sentence(s) # in a chapter;
6. When are other suspects first introduced - chapter #, the sentence(s) # in a chapter.

Based on our analysis, we obtained the following results:

3.1 The Middle of Things

Total Chapters: 29

word	context word	count
Cortelyon	Dr	10
Cortelyon	Ashton	5
Cortelyon	hed	3
Cortelyon	asked	2
Cortelyon	square	2
Cortelyon	years	2
Cortelyon	Cave	2
Cortelyon	wed	2
Cortelyon	papers	2
Cortelyon	found	2

Table 1: Caption

1. Detective Viner is first referenced in chapter 1 sentence 1.

2. Perpetrator Cortelyon first referenced chapter 2 sentence 54. The text reads, "Cortelyon saw that by killing Ashton he alone would have the secret; he evidently got two accomplices who were necessary to him."

3. The crime of murder first appears in chapter 1 sentence 76. The text reads, "Then he saw white linen, and a bloodstain slowly spreading over its glossy surface."

4. There are 90 context words total. See Table 1

5. Detective-perp cooccurrences in (chapter, sentence) form: [(3, 78), (28, 134)]

6. Other Suspects Penkridge first referenced chapter 1 sentence 1

3.2 Scarhaven Keep

Total Chapters: 31

1. Detective Richard Copplestone first referenced chapter 1 sentence 39

2. There are a series of perpetrators. Addie Chatfield and her husband impersonate a rich person. With the help of her father, Peter, they steal the Greyle's family weath. They are first referenced in chapter 4 sentence 114.

3. There is a plot twist introduced in the last chapter. Originally they are investigating the murder of Basset Oliver. This turns out to be an accident not murder. You then find out the real crime is stealing money.

4. The number of context words is 969. See Table 2

5. Detective-perp cooccurrences in (chapter, sentence) form: [(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. Audrey Greyle is a suspect at first. She is introduced chapter 3 sentence 166.

3.3 The Talleyrand Maxim

Total Chapters: 28

word	context word	count
Chatfield	Mr	74
Chatfield	Peter	46
Chatfield	said	34
Chatfield	Addie	30
Chatfield	Miss	30
Addie	Chatfield	30
Chatfield	I	26
Chatfield	Vickers	26
Peter	Chatfield	23
Chatfield	know	18
Chatfield	replied	16

Table 2: Caption

1. Detective Linford Pratt first referenced chapter 1 sentence 1.

2. The perpetrator Harper Mallathorpe is first referenced chapter 1 sentence 57.

3. Crime murder first referenced chapter 7 sentence 93. There is a plot twist in the last chapter. Initially they investigate a murder, but it turns out the man left town and went to America. The real crime is that Harper destroyed her husband's will because he left his money to a charity. You can see this plot twist visually by inspecting Figure 1. John Mallathrope's will is introduced early on as the murder is investigated. Then he leave it alone for a few chapters. Fletcher reintroduces it again a few chapters before the end. Then in the last chapter the plot twist occurs, and he ties it all together.

4. The number of context words is 635. See Table 3.

5. Detective-perp cooccurrences in (chapter, sentence) form: [(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. Other Suspects Nesta first referenced chapter 2 sentence 17.

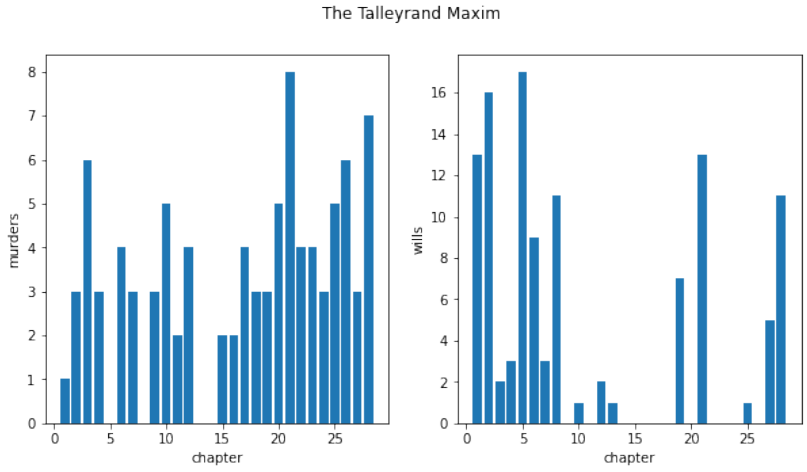

Figure 1: The Talleyrand Maxim

word	context word	count
Mallathorpe	Mrs	388
Mallathorpe	Miss	104
Mallathorpe	John	80
Mallathorpe	Nesta	60
Mallathorpe	Mr	56
Mallathorpe	I	48
Mallathorpe	Pratt	44
Mallathorpe	said	40
Mallathorpe	Harper	30
Harper	Mallathorpe	30
Mallathorpe	see	28
Mallathorpe	know	28
Mallathorpe	would	18
Mallathorpe	Eldrick	18
Mallathorpe	time	16
Mallathorpe	Normandale	16
Mallathorpe	letter	16
Mallathorpe	called	16
Mallathorpe	went	16
Mallathorpe	doesnt	16
Mallathorpe	late	14
Mallathorpe	family	14
Mallathorpe	But	14
Mallathorpe	business	14
Mallathorpe	mother	14
Mallathorpe	left	12
Mallathorpe	property	12
Mallathorpe	died	12

Table 3: Caption

3.4 The Middle Temple Murder

Total Chapters: 36

1. Detective Frank Spargo is first referenced in chapter 1 sentence 1.
2. Perpetrator Jane Baylis first referenced chapter 19 sentence 24.
3. The crime of murder first appears in referenced chapter 1 sentence 24.
4. There are 360 context words total. See Table 4
5. Detective-perp cooccurrences in (chapter, sentence) form: [(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. Other Suspects Chamberlayne first referenced chapter 6 sentence 156.

3.5 Dead Men's Money

Total Chapters: 37

1. Detective Hugh Moneylaw is first referenced in chapter 1 sentence 19.

word	context word	count
Baylis	Miss	96
Baylis	Jane	24
Baylis	But	14
Baylis	Spargo	14
Jane	Baylis	12
Baylis	see	8
Baylis	said	8
Baylis	I	8
Baylis	told	8
Baylis	looked	6
Baylis	scornful	6
Baylis	suddenly	6
Baylis	came	6

Table 4: Caption

word	context word	count
Gilbert	Sir	156
Gilbert	Carstairs	91
Carstairs	Gilbert	91
Carstairs	Sir	84
Gilbert	I	30
Carstairs	Michael	27
Carstairs	Lady	15
Carstairs	Mr	14
Gilbert	Mr	12
Carstairs	I	12
Gilbert	man	11
Carstairs	yacht	8
Carstairs	Carstairs	8
Gilbert	Hathercleugh	7

Table 5: Caption

2. Perpetrator Gilbert Carstairs first referenced chapter 7 sentence 6.

3. The crime of murder first appears in referenced chapter 1 sentence 3.

4. There are 676 context words total. See Table 5

5. Detective-perp cooccurrences in (chapter,sentence) form: [(22, 67)]

6. Dead Men's Money has no Other Suspects.

4 Summary

In the research we have presented here, plot summaries were crafted on five select works from author Joseph Smith Fletcher. Regular expressions (RegEx were then utilized to perform computational analysis of these five novels. As part of the data preparation, NLTK was used to prevent false negatives in the RegEx matching, and WordNet was utilized to help identify possible synonyms of key target words, such as "will" or "murder". Additionally, chapters were tokenized into sentences using a custom defined sentence-splitting regex function. This was then coupled with NLTK's Treebank Word Tokenizer to produce the dataset that we used for the analysis. Similarly, the analysis was performed using regex operations and the results in section 3 show the data points used to answer each question for each book. Additionally, figures 1-5 show the frequency of words "murder" & "will" across the chapters for all five books. Generally, this shows that there is some inverse relationship between mentions of "murder" and "will" (I.E. "murder" will be heavily

Table 6: Information of "The Middle of Things"

The Middle of Things	Detective	'Viner'
	Perpetrator	'Cortelyon'
	Crime	'murder'
	Other Suspects	'Penkridge'
	Victims	'Richard Ashton'
	Characters	'Viner', 'Pawle','Ashton', 'Carless','Wickham','Hyde','Marketstoke', 'Ellingham','Drillford','Killenhall','Methley','Perkwite','Miss Wickham', 'Felpham','John Ashton','Cave','Miss Penkridge','Fosdick', 'Millington-Bywater','Cortelyon','Armitstead','Barleyfield','Summers', 'Van Hoeren','Millwaters','Woodlesford','Penkridge','Stephens', 'Portlethwaite','Martincole','Richard','Bigglesforth','Driver', 'Nugent Starr','Avice','Pelver','Earl','Langton Hyde','Roscombe', 'Langton','Lincoln'

Table 7: Information of "The Middle Temple Murder"

The Middle Temple Murder	Detective	"Spargo | Frank Spargo | Frank"
	Perpetrator	"Jane Baylis | Miss Baylis"
	Crime	"murder"
	Other Suspects	"Stephen Aylmore | Chamberlayne | Myerst"
	Victims	"John Maitland | John Marbury"
	Characters	'Spargo','Breton','Marbury','Elphick','Aylmore','Maitland', 'Rathbury','Quarterpage','Cardlestone','Myerst','Baylis', 'John Marbury','John Maitland','Ronald Breton', 'Walters','Criedir', 'Gutch','Driscoll','Mother Gutch', 'Webster','Anderson', 'Jane Baylis','Mollison','Jessie Aylmore','Starkey','Jessie','Stephen Aylmore','Evelyn', 'Cooper','Ronald','Crowfoot','Doolittle', 'Ainsworth','David Lyell','Vallas','Kaye','Stephens','Septimus Elphick','Miss Baylis','Robertson','Lyell'

Table 8: Information of "Scarhaven Keep"

Scarhaven Keep	Detective	"Copplestone | Richard Copplestone"
	Perpetrator	"Martin, Addie Chatfield"
	Crime	"fraud"
	Other Suspects	"Audrey Greyle"
	Victims	"Audrey Greyle"
	Characters	'Copplestone','Greyle','Gilling','Chatfield','Vickers','Audrey','Marston Greyle','Bassett Oliver','Cresswell','Oliver','Stafford', 'Petherton', 'Wooler','Addie', 'Dennie','Cresswell Oliver','Peter Chatfield','Spurge', 'Andrius','Pike','Ewbank','Addie Chatfield','Zachary Spurge','Scarhaven', 'Squire','Jim','Peter','Altmore','Audrey Greyle','Bassett','Martin', 'Rothwell','Montmorency','Hackett','Miss Greyle','Glen','Salmon', 'Jerramy','Elkin','Hobkin','Tretheway','Valdey','Jim Spurge','Valentine Greyle','Stephen John'

Table 9: Information of "The Talleyrand Maxim"

The Talleyrand Maxim	Detective	"Linford Pratt | Pratt"
	Perpetrator	"Mallathorpe | Harper | Harper Mallathorpe"
	Crime	"murder"
	Other Suspects	"Nesta"
	Victims	"John Mallathorpe"
	Characters	'Pratt','Eldrick','Collingwood','Mallathorpe','Nesta','Parrawhite', 'Murgatroyd','John Mallathorpe','Pickard','Byner','Esther Mawson','Bartle','Antony Bartle', 'Cobcroft','Prydale','Nesta Mallathorpe','Harper','Robson','Normandale Grange','Esther', 'Linford Pratt','Miss Mallathorpe','Harper Mallathorpe', 'Gaukrodger','James Parrawhite','Marshall','Barford','Pascoe', 'Jabey Naylor','Parsons','Clough','Stringer','Shepherd', 'Bartle Collingwood'

Table 10: Information of "Dead Men's Money"

Dead Men's Money	Detective	Hugh Moneylaw | Hugh | Moneylaw
	Perpetrator	"Gilbert Carstairs | Meekin"
	Crime	"murder"
	Other Suspects	"Chamberlayne"
	Victims	"Crone"
	Characters	'Lindsey','Gilbert','Portlethorpe','Chisholm','Gilverthwaite','Gilbert Carstairs', 'Crone','Phillips','Maisie','Smeaton','Murray','Hugh','Hollins','Moneylaws', 'Michael Carstairs','Carstairs','Ridley','Ralston','Gavin Smeaton','John Phillips','James Gilverthwaite','Lady Carstairs','Peebles','Elphinstone', 'Andrew Dunlop','Abel Crone','Michael','Paley','Hanson','Carter','Alexander', 'Maisie Dunlop','Meekin','Nance Maguire','Hathercleugh','Dunlop','John Paley','Robertson','Tom Dunlop','Hugh Moneylaws','Scott','Watson','Martin Smeaton','Turndale'

mentioned in one chapter then immediately decrease in the next chapters while "will" is sparsely mentioned in the earlier chapter and increases as "murder" decreases).

References

[1] J. Fletcher, *Dead Men's Money*. 1920.

[2] J. Fletcher, *The Middle of Things*. 1900.

[3] J. Fletcher, *The Middle Temple Murder*. 1919.

[4] J. Fletcher, *The Tallyrand Maxim*. 1919.

[5] J. Fletcher, *Scarhaven Keep*. 1920.

[6] "Project gutenberg." `gutenberg.org`.

[7] "Goodreads." `www.goodreads.com`.

[8] "Kadaxis." `https://kadaxis.com/book/`.

A Sentence Splitting Regex

Choosing when and how to split a chapter into sentences is a hard task. In order to do so, some decisions need to be made. In Figure 2 is the regex we created and an explanation of both it and the cases it handles.

```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)
```

Figure 2: A short description of each part of the regex, the cases, and examples of how it splits sentences. Included as a screenshot of python code due to lack of time.