

Cyfrowe narzędzia w przekładoznawstwie

ALEKSANDRA RYKOWSKA

ALEKSANDRA.RYKOWSKA@DOCTORAL.UJ.EDU.PL

Rolling stylometry

- The Rolling Stylometry method is designed to detect **stylistic takeovers**.
- To give an example: imagine a text that is supposedly written by more than one author, and you want to pinpoint the stylistic change between the individual authorial “signals”. In the case of the earliest Polish translation of the Bible knowns as Queen Sophia’s Bible, one might be interested in tracing the five scribal hands evidenced in the manuscript – were they “just” the scribes, or were they also involved in actual translatorial? In Fig. 1 above, one can notice that indeed, the change of a scribe (vertical dashed lines) is usually correlated by stylistic change (different colors of the strip).

Rolling stylometry

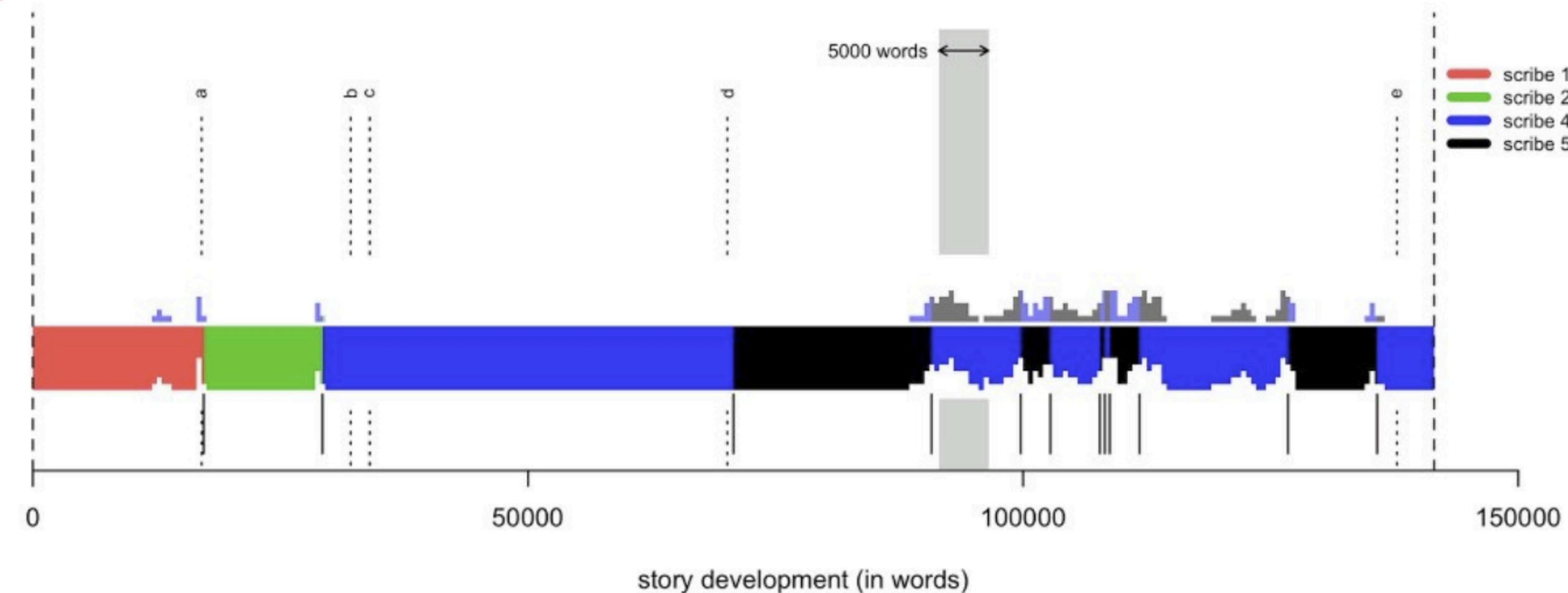
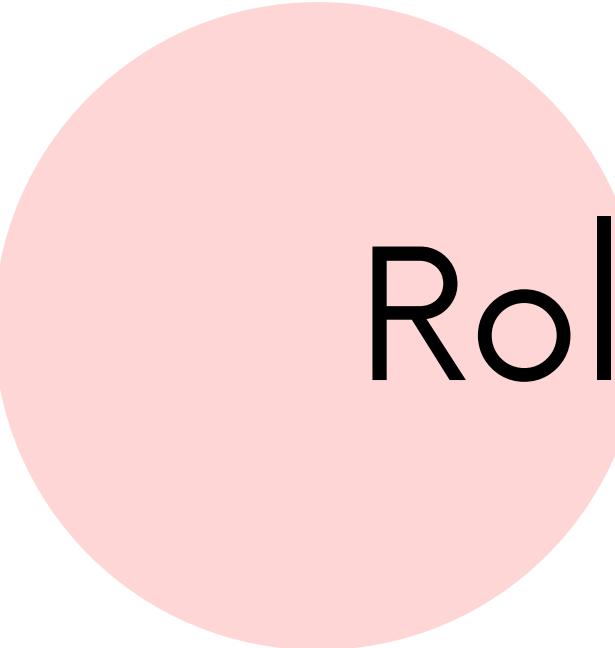


Fig. 1: *Queen Sopia's Bible* examined using Rolling Stylometry (in SVM flavor), with the following parameters: 100 MFWs, window size of 5,000 words, sample overlap of 4,500 words.



Rolling stylometry

- Tekst jest dzielony na równe części
- Zamiast porównywać cały tekst z innymi tekstami, porównujemy te pocięte części pod względem podobieństwa
- “Arguably, any classification method can be combined with this procedure. However, so far, the method implementation in *stylo* includes support vector machines (SVM), nearest shrunken centroids (NSC), and Delta in its classical Burrowsian flavor”. ([Eder 2016](#))

Rolling stylometry

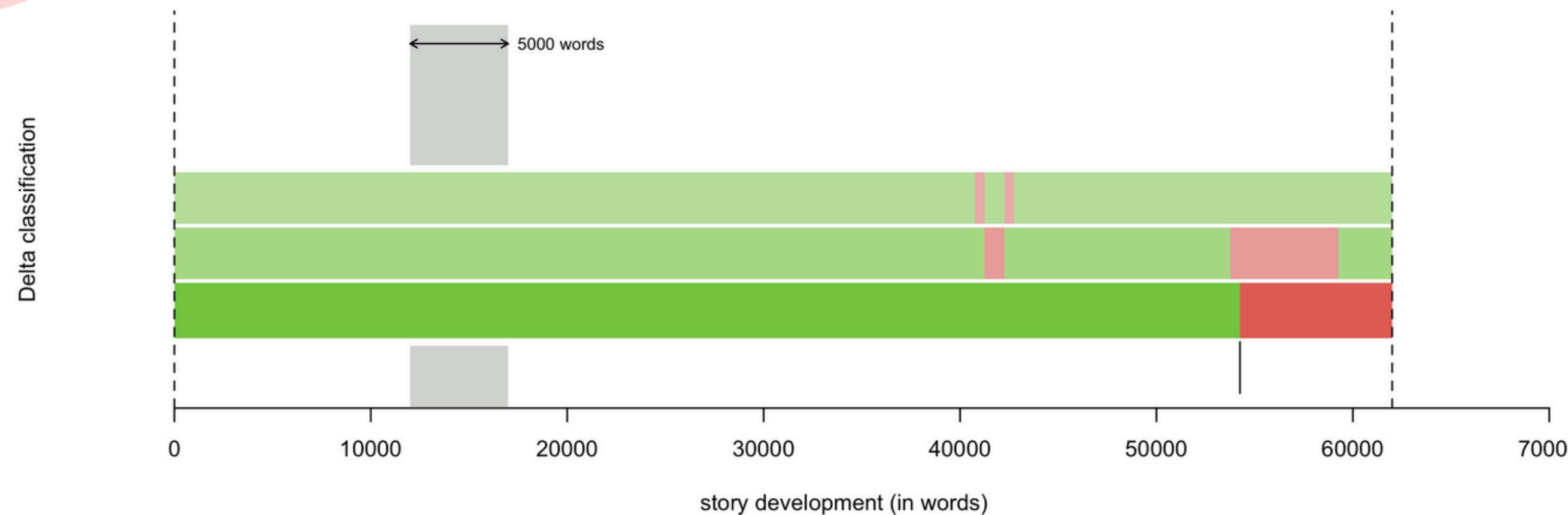


Fig. 3: The Inheritors by Conrad/Ford assessed using Rolling Delta and 1,000 MFWs. The bottom stripe indicates the first ranked candidate (i.e. the most probable), then comes the second and the third suggested class. Sections attributed to Ford are marked green, the red ones are for Conrad.

Rolling stylometry

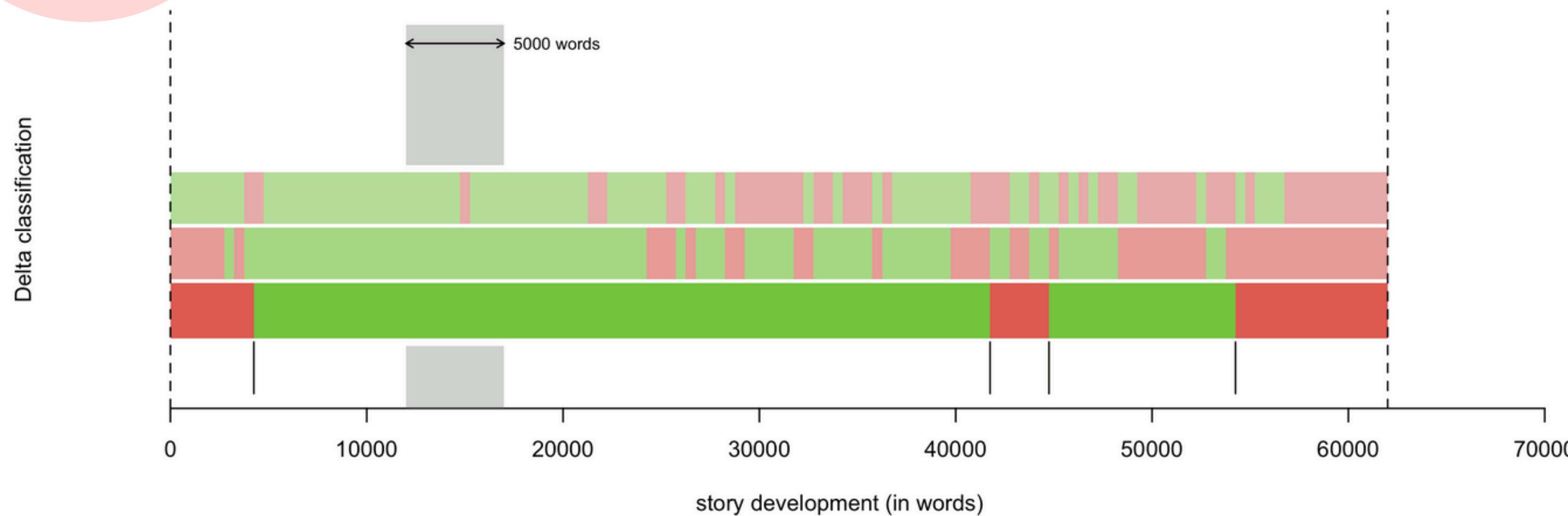


Fig. 4: *The Inheritors* by Conrad/Ford assessed using Rolling Delta and 500 MFWs.

Rolling stylometry - struktura korpusu

Nasz korpus musi obowiązkowo zawierać 2 foldery o podanych nazwach (uwaga na małe litery!)

test_set

W folderze ma znaleźć się tekst, którego autorstwo chcemy ustalić

reference_set

W folderze mają znaleźć się teksty autorów, których chcemy “wykryć”

Rolling stylometry - struktura korpusu

Jeśli foldery nie będą tak zatytułowane (albo będą literówki, wielkie litery), to pokaże nam się błąd:

test_set

W folderze ma znaleźć się tekst, którego autorstwo chcemy ustalić

reference_set

W folderze mają znaleźć się teksty autorów, których chcemy "wykryć"

```
!!!!!!  
Working directory should contain two subdirectories:  
    "reference_set" and "test_set"  
!!!!!!
```

Rolling stylometry - funkcja

rolling.classify()

UWAGA: ta funkcja nie ma interfejsu, więc wywołana w ten sposób automatycznie zrobi analizę za pomocą klasycznej delty

Rolling stylometry - metody

SVM

Support Vector Machine

NSC

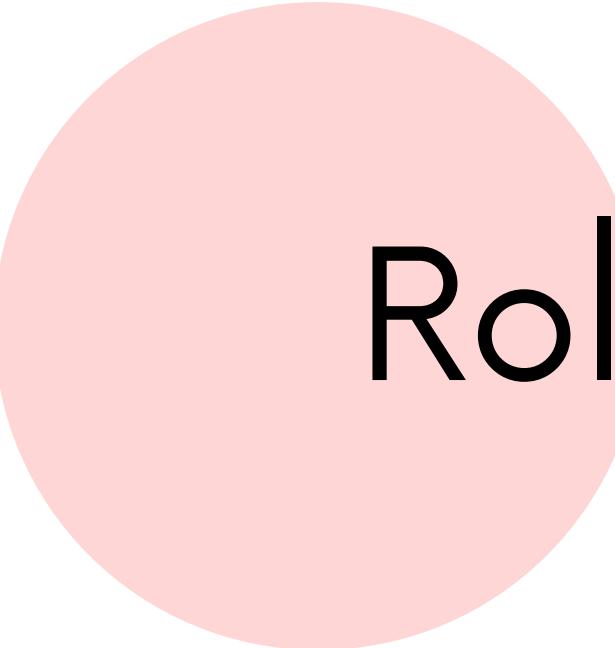
Nearest Shrunken Centroid

Delta

Jeśli będziemy robić tego typu analizy, to dobrze przetestować wszystkie i porównać wyniki 😊

Zanim rolling stylo

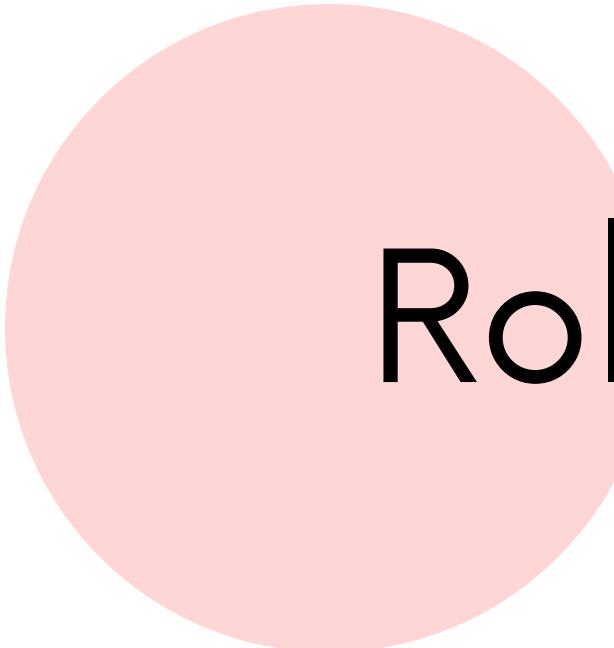
- Pamiętamy o wczytaniu biblioteki
library(stylo)
 - Pamiętamy o ustawieniu ścieżki do naszego folderu
setwd("ściezka/do/naszego/folderu")
albo na Windowsie: Plik → Zmień katalog
 - (opcjonalnie, przy używaniu setwd polecane:) Sprawdzamy, czy na pewno jesteśmy w dobrym folderze
getwd()



Rolling stylometry

- Jeżeli chcemy doczytać więcej o funkcji, wpisujemy w konsoli:
help(rolling.classify)

Przeniesie nas to automatycznie na stronę dokumentacji:
<http://127.0.0.1:11310/library/stylo/html/rolling.classify.html>



Rolling stylometry

- Jeżeli chcemy doczytać więcej o funkcji, wpisujemy w konsoli:
help(rolling.classify)
- Disclaimer: w dokumentacji jest informacja, że po wywołaniu funkcji z poleceniem “gui”, będziemy mogli skorzystać z interfejsu użytkownika, ale niestety nie jest jeszcze opracowany:

```
> rolling.classify(gui = TRUE)
using current directory...

GUI could not be launched -- it is not supported yet :-(
```

Rolling stylometry - metody

SVM

Support Vector Machine

NSC

Nearest Shrunken Centroid

Delta

```
rolling.classify(write.png.file = TRUE, classification.method = "svm",  
mfw=100, training.set.sampling = "normal.sampling", slice.size = 5000,  
slice.overlap = 4500)
```

```
rolling.classify(write.png.file = TRUE, classification.method = "svm",  
mfw=100, training.set.sampling = "normal.sampling", slice.size = 5000,  
slice.overlap = 4500)
```

- **rolling.classify()** → nasza funkcja, w nawiasie dodajemy parametry, które chcemy wziąć pod uwagę w analizie
 - **write.png.file = TRUE** → wykres zapisze się nam w formacie .png w folderze
 - **classification.method = “metoda”** → ustawiamy jedną z trzech wymienionych metod
 - **mfw = 100** → analiza będzie brała pod uwagę 100 najczęściej występujących słów
 - **training.set.sampling** → umożliwia podzielenie tekstów, których autorstwo jest znane w pełni na mniejsze części, mamy do wyboru opcje no.sampling (defaultowe), normal.sampling, random.sampling
 - **slice.size** → umożliwia ustawienie liczby słów, która wejdzie w nasze “okienko” w analizie
 - **slice.overlap** → przy slice.size=5000, slice.overlap=4500 oznacza, że pierwsza próbka weźmie pod uwagę słowa 1–5,000, kolejna 501–5,500, następna 1001–6,000, itd.

Inne przydatne parametry

`milestone.points`

sometimes, there is a need to mark one or more passages in an analyzed text (e.g. when external evidence suggests an authorial takeover at a certain point) to compare if the *a priori* knowledge is confirmed by stylometric evidence. To this end, one should add into the test file a string "xmilestone" (when input texts are loaded directly from files), or specify the break points using this parameter. E.g., to add two lines at 10,000 words and 15,000 words, use `milestone.points = c(10000, 15000)`.

`milestone.labels`

when milestone points are used (see immediately above), they are automatically labelled using lowercase letters: "a", "b", "c" etc. However, one can replace them with custom labels, e.g. `milestone.labels = c("Act I", "Act II")`.

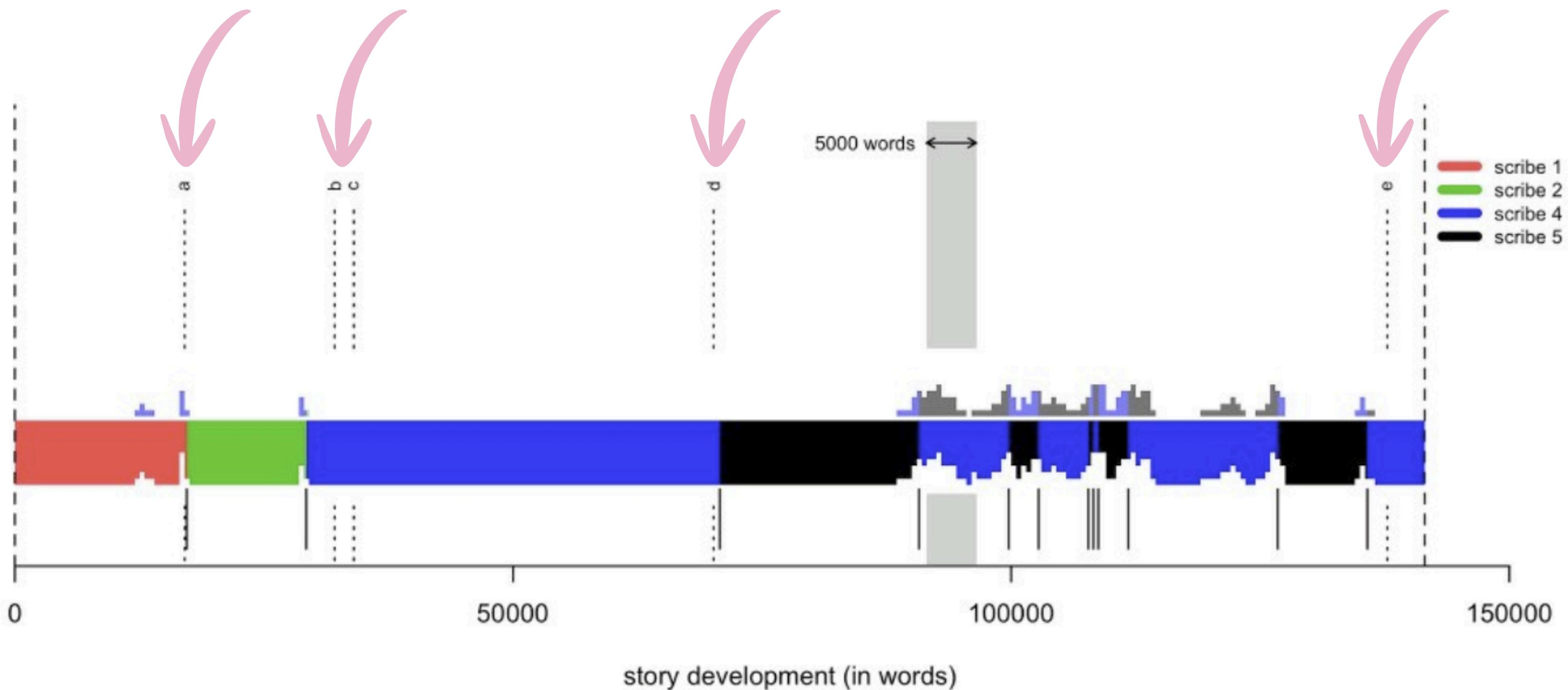


Fig. 1: *Queen Sopia's Bible* examined using Rolling Stylometry (in SVM flavor), with the following parameters: 100 MFWs, window size of 5,000 words, sample overlap of 4,500 words.

Rolling stylometry - metody

SVM

Support Vector Machine

NSC

Nearest Shrunken Centroid

Delta

```
rolling.classify(write.png.file = TRUE, classification.method = "nsc",  
mfw=100, training.set.sampling = "normal.sampling", slice.size = 5000,  
slice.overlap = 4500)
```

Rolling stylometry - metody

SVM

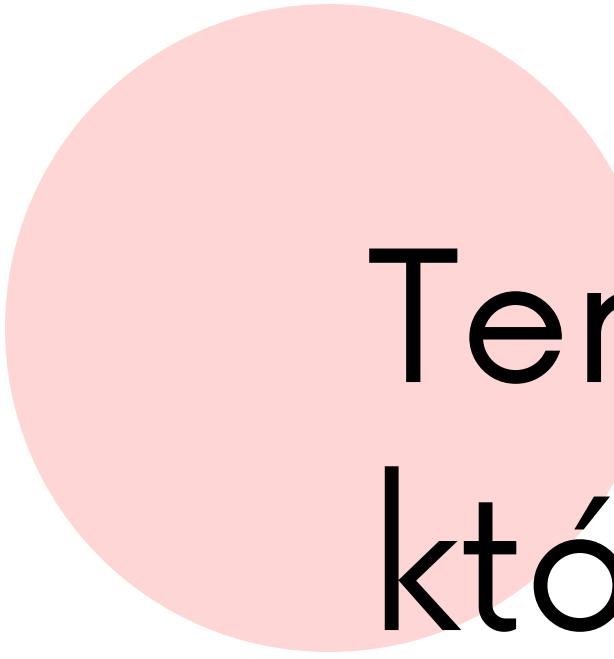
Support Vector Machine

NSC

Nearest Shrunken Centroid

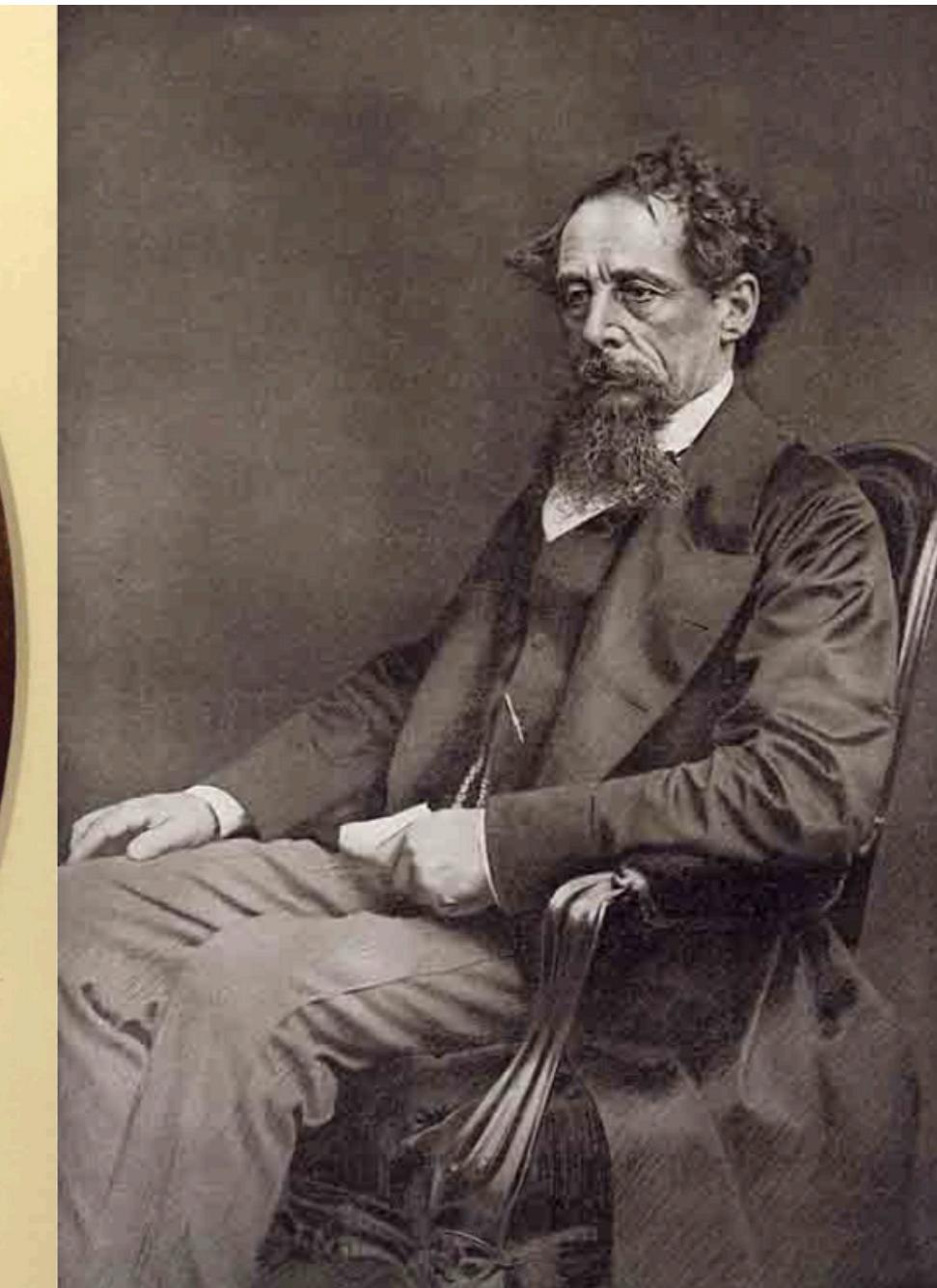
Delta

```
rolling.classify(write.png.file = TRUE, classification.method =  
"delta", mfw=500)
```



Teraz pozostaje znaleźć utwory,
które były pisane lub tłumaczone
przez więcej niż jedną osobę

Dickens and Collins



No Thoroughfare is a [stage play](#) and novel by [Charles Dickens](#) and [Wilkie Collins](#), both released in December 1867.

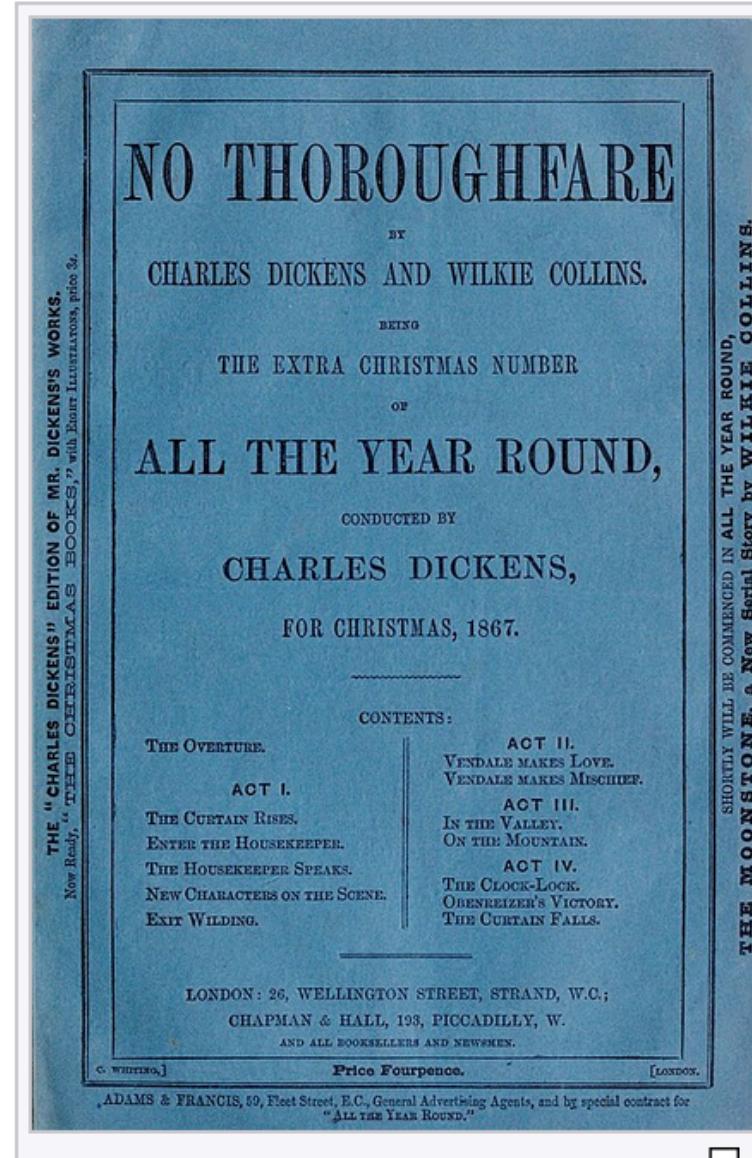
Background [edit]

In 1867 [Charles Dickens](#) and [Wilkie Collins](#) collaborated to produce a [stage play](#) titled *No Thoroughfare: A Drama: In Five Acts*.^[1] The two had previously collaborated on the play *The Frozen Deep*. This was the last stage production to be associated with Dickens, who died in June 1870. The play opened at the [Adelphi Theatre](#) on 26 December 1867.

The novel *No Thoroughfare* was also first published in 1867, in the [Christmas](#) number of Dickens's periodical *All the Year Round*. There are thematic parallels with other books from Dickens's mature writings, including *Little Dorrit* (1857) and especially *Our Mutual Friend* (1865).

The publication of the story in *All The Year Round* represents an early example of commercial [merchandising](#), promoting the story to those who were aware of the stage play, and the play to those who had read the book. The chapters of the book are referred to as 'acts', and match the acts of the play.

In the book Collins assisted in Act 1 and Act 4;^[2] Collins scripted most of the stage play with Dickens's assistance.^[3]



First edition cover

No Thoroughfare is a stage play and novel by Charles Dickens and Wilkie Collins, both released in December 1867.

Background [edit]

In 1867 Charles Dickens and Wilkie Collins collaborated to produce a stage play titled *No Thoroughfare: A Drama: In Five Acts*.^[1] The two had previously collaborated on the play *The Frozen Deep*. This was the last stage production to be associated with Dickens, who died in June 1870. The play opened at the Adelphi Theatre on 26 December 1867.

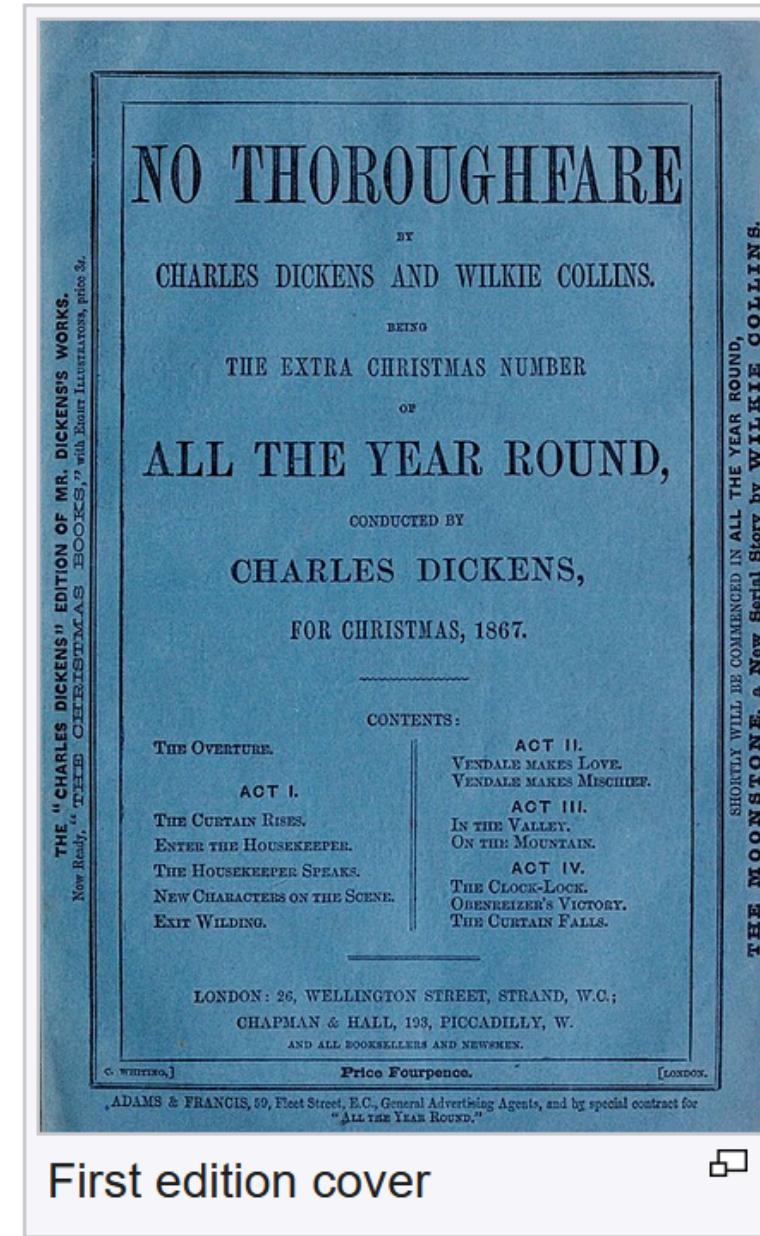
The novel *No Thoroughfare* was also first published in 1867, in the Christmas number of Dickens's periodical *All the Year Round*. There are thematic parallels with other books from Dickens's mature writings, including *Little Dorrit* (1857) and especially *Our Mutual Friend* (1865).

The publication of the story in *All the Year Round* represents an early example of commercial merchandising, promoting the story to those who were aware of the stage play, and the play to those who had read the book.

The chapters of the book are referred to as 'act'

Sprawdźmy to :)

In the book Collins assisted in Act 1 and Act 4;^[2] Collins scripted most of the stage play with Dickens's assistance.^[3]



First edition cover

Zadanie

- Używając funkcji `rolling.classify()`, sprawdź, jakie fragmenty dramatu *No throughfare* najprawdopodobniej napisał Dickens, a które Collins.
- Użyj wszystkich trzech metod analizy
 - SVM ($mfw = 100$)
 - NSC ($mfw = 100$)
 - Delta ($mfw = 300$)

Projekt zaliczeniowy

Czy mam pomysł na temat?

- Jeśli tak: czy potrzebuję go uściślić?
- Jeśli nie: czy wiem, w jakim obszarze chcę szukać tematu?

Czy wiem, jakich tekstów szukać? Czy wiem, ile tekstów mi będzie potrzebne?

Czy już teraz wiem, że będę z czymś miał(a) problem przy temacie, który chcę realizować?

Do zobaczenia za tydzień!

