

The Adventures of Sherlock Holmes

Total words in novel : 104,410

Total unique words in novel : 8992

Most frequent words:



1. The 5248
2. I 2993
3. And 2800
4. To 2654
5. Of 1683
6. A 2521
7. In 1683
8. That 1641
9. Was 1390
10. It 1273
11. You 1214
12. He 1155
13. Is 1105
14. His 1103
15. My 906
16. Have 902
17. Had 820
18. With 802
19. As 769
20. Which 761

Most interesting frequent words:

Without Lemmatization :

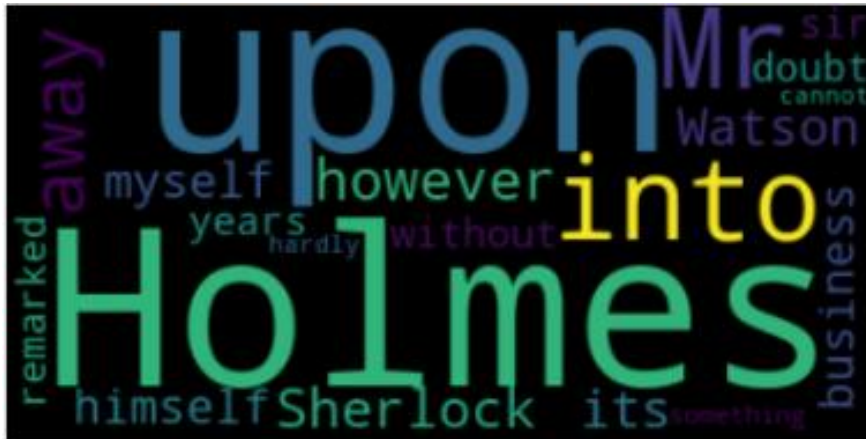


The list of common words does not include past tenses or plural words like asked , cried, seemed, eyes but their present and singular version are present .

So, to get the most interesting frequent words , we can use NLTK library to do lemmatization or having another exhaustive common word list.

1. Upon 462
2. Holmes 457
3. Into 275
4. Mr 271
5. Away 104
6. However 98
7. Sherlock 94
8. Its 89
9. Asked 988
10. Eyes 87
11. Watson 81
12. Myself 77
13. Himself 73
14. Business 73
15. Seen 72
16. Hands 72
17. Without 71
18. Cried 69
19. Seemed 68
20. Sir 67

With Lemmatization:



20 Least Frequent words:



1. Walsall 1
2. Manifested 1
3. Violet 1
4. Mauritius 1
5. Solely 1
6. Survived 1
7. Stand 1
8. Locus 1
9. Freehanded 1
10. Kindspoken 1
11. Arguments 1
12. Blockaded 1
13. Seaman 1
14. Persevering 1
15. Brainfever 1
16. Slighted 1

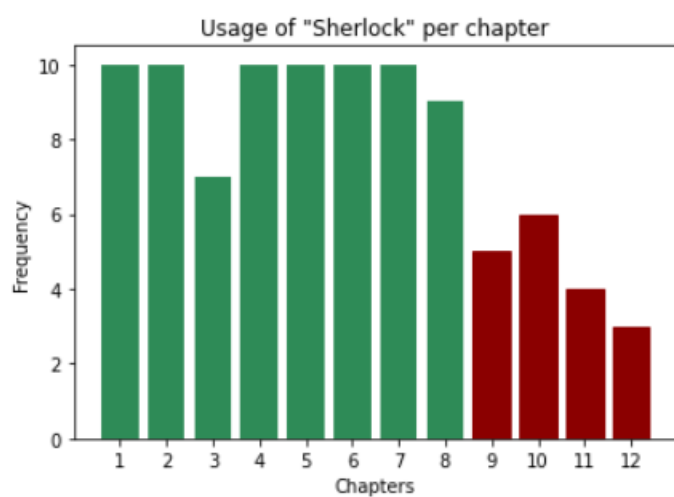
- 17. Alices 1
- 18. Assembled 1
- 19. Relieve 1
- 20. Sobered 1

Interesting word usage pattern:

The word Sherlock is interesting in the sense that it was used a very frequently in the beginning, and during the middle, and then the frequency decreases towards the end of the book.

Number of usages in each chapter of book

[10, 10, 7, 10, 10, 10, 10, 9, 5, 6, 4, 3]



Favorite quote:

“There is nothing more deceptive than an obvious fact” - Chapter 4

Generated sentence:

“Sherlock Holmes gravely Do you a black lace which you managed to a waste the upper lip to say nothing”

Sentence Autocomplete:

```
print(p.getAutocompleteSentence('I have no do'))
```

```
['I have no do', 'I have no doubt', 'I have no doubts', 'I have no doubting', 'I have no doubtless', 'I have no doubted', 'I have no double', 'I have no doublebreasted', 'I have no doublebedded', 'I have no doubleedged', 'I have no doubled', 'I have no doubly', 'I have no doing', 'I have no doings', 'I have no door', 'I have no doormat', 'I have no doors', 'I have no doorstep', 'I have no doorway', 'I have no doorthat', 'I have no dooties', 'I have no down', 'I have no downward', 'I have no downstairs', 'I have no downry', 'I have no dont', 'I have no dontits', 'I have no donna', 'I have no done', 'I have no dock', 'I have no docketing', 'I have no docks', 'I have no doctor', 'I have no doctors', 'I have no document', 'I have no does', 'I have no doesnt', 'I have no dozen', 'I have no dog', 'I have no dogcart', 'I have no dogwhip', 'I have no doreally', 'I have no doddering', 'I have no dollars', 'I have no dottles']
```

The autocomplete works by creating a Trie of all the words from the novel and given a sentence , it looks at the last incomplete word and try to predict all possible words.

If given a sentence , it does not guarantee that the entire sentence is part of the novel but only the current word for whom it is autocompleting. That is why it is possible that the entire sentence does not make any sense.

In the example given above, the function gives all possible words staring from Prefix "Do" .