

```

# !/usr/bin/env python
# (c) hughes
# -----
#
#

# ----- READS THE BOOK INTO A STRING -----

import sys
import re

def readbook(Alice, clean=True):
    with open('Alice.txt', 'r') as myfile:
        data = myfile.read().replace('\n', ' ').replace('\r', ' ').replace('\t', ' ')
    if clean:
        data = data.replace('; ', ' ').replace(', ', ' ').replace(':', ' ')
        data = data.replace('-', ' ').replace('"', ' ').replace("'", " ").lower()
    return data

def get_sentence(data):
    sentence = data.split(" ")
    numSentences = sum(len(word) for word in sentence) / len(sentence)
    return numSentences

# count words in sentence
def count_words(numSentences):
    numWords = 0
    for sentence in numSentences:
        numSentences = sentence.split(' ')
        numWords += len(numSentences)/len()
    return numWords

# count chars per words
def count_chars_in_word(numWords):
    numChars = 0
    for sentence in numWords:
        numWords = sentence.split(' ')
        numChars += list(numWords)
    return numChars

wordcount = {}
for words in sentence:
    if words in wordcount:
        wordcount[words] += 1
    else:
        wordcount[words] = 1
frequency = sorted(wordcount, key = wordcount.get, reverse = True)
top100 = frequency[:100]

```

```

if __name__ == "__main__":
    print("Alice in Wonderland")
    data = readbook("Alice.txt")
    print("Sentences: ", (numSentences))
    print("Words: ", (numWords))
    print("Characters: ", (numChars))
    print("Top 100 most common words: ", top100)

    Alice in Wonderland
    Sentences:  3.759879563417388
    Words:  31271
    Characters:  164410
    Top 100 most common words:  [' ', 'the', 'and', 'to', 'a', 'of', 'she', 'said', '

```

Double-click (or enter) to edit

```

import sys
import re

def readbook(Strange, clean=True):
    with open('The_Strange_Case.txt', 'r') as myfile:
        data = myfile.read().replace('\n', ' ').replace('\r', ' ').replace('\t', '')
    if clean:
        data = data.replace('; ', ' ').replace(', ', ' ').replace(':', ' ')
        data = data.replace('-', ' ').replace('"', ' ').replace("'", " ").lower()
    return data

def get_sentence(data):
    sentence = data.split(" ")
    numSentences = sum(len(word) for word in sentence) / len(sentence)
    return numSentences

# count words in sentence
def count_words(numSentences):
    numWords = 0
    for sentence in numSentences:
        numSentences = sentence.split(' ')
        numWords += len(numSentences)/len()
    return numWords

# count chars per words
def count_chars_in_word(numWords):
    numChars = 0
    for sentence in numWords:
        numWords = sentence.split(' ')
        numChars += list(numWords)
    return numChars

wordcount = {}
for words in sentence:

```

```
if words in wordcount:
    wordcount[words] += 1
else:
    wordcount[words] = 1
frequency = sorted(wordcount, key = wordcount.get, reverse = True)
top100 = frequency[:100]
```

```
if __name__ == "__main__":
    print("The Strange Case")
    data = readbook("The_Strange_Case.txt")
    print("Sentences: ", (numSentences))
    print("Words: ", (numWords))
    print("Characters: ", (numChars))
    print("Top 100 most common words: ", top100)
```

```
The Strange Case
Sentences:  3.759879563417388
Words:  31271
Characters:  164410
Top 100 most common words:  ['', 'the', 'and', 'to', 'a', 'of', 'she', 'said', ']
```