

## Question 1

```
In [4]: from nltk.book import *
*** Introductory Examples for the NLTK Book ***
Loading text1, ..., text9 and sent1, ..., sent9
Type the name of the text or sentence to view it.
Type: 'texts()' or 'sents()' to list the materials.
text1: Moby Dick by Herman Melville 1851
text2: Sense and Sensibility by Jane Austen 1811
text3: The Book of Genesis
text4: Inaugural Address Corpus
text5: Chat Corpus
text6: Monty Python and the Holy Grail
text7: Wall Street Journal
text8: Personals Corpus
text9: The Man Who Was Thursday by G . K . Chesterton 1908
```

### SAMPLE EXERCISE

```
In [12]: #Using Concordance
text1.concordance("monstrous")
print("-"*100)
text2.concordance("affection")
print("-"*100)
text3.concordance("lived")
```

Displaying 11 of 11 matches:

ong the former , one was of a most monstrous size . ... This came towards us ,  
ON OF THE PSALMS . " Touching that monstrous bulk of the whale or ork we have r  
ll over with a heathenish array of monstrous clubs and spears . Some were thick  
d as you gazed , and wondered what monstrous cannibal and savage could ever hav  
that has survived the flood ; most monstrous and most mountainous ! That Himmel  
they might scout at Moby Dick as a monstrous fable , or still worse and more de  
th of Radney .'" CHAPTER 55 Of the Monstrous Pictures of Whales . I shall ere l  
ing Scenes . In connexion with the monstrous pictures of whales , I am strongly  
ere to enter upon those still more monstrous stories of them which are to be fo  
ght have been rummaged out of this monstrous cabinet there is no telling . But  
of Whale - Bones ; for Whales of a monstrous size are oftentimes cast up dead u

---

-----  
-----  
Displaying 25 of 79 matches:

, however , and , as a mark of his affection for the three girls , he left them  
t . It was very well known that no affection was ever supposed to exist between  
deration of politeness or maternal affection on the side of the former , the tw  
d the suspicion -- the hope of his affection for me may warrant , without impru  
hich forbade the indulgence of his affection . She knew that his mother neither  
rd she gave one with still greater affection . Though her late conversation wit  
can never hope to feel or inspire affection again , and if her home be uncomfo  
m of the sense , elegance , mutual affection , and domestic comfort of the fami  
, and which recommended him to her affection beyond every thing else . His soci  
ween the parties might forward the affection of Mr . Willoughby , an equally st  
the most pointed assurance of her affection . Elinor could not be surprised at  
he natural consequence of a strong affection in a young and ardent mind . This  
opinion . But by an appeal to her affection for her mother , by representing t  
every alteration of a place which affection had established as perfect with hi  
e will always have one claim of my affection , which no other can possibly shar  
f the evening declared at once his affection and happiness . " Shall we see you  
ause he took leave of us with less affection than his usual behaviour has shewn  
ness ." " I want no proof of their affection , " said Elinor ; " but of their en  
onths , without telling her of his affection ;-- that they should part without  
ould be the natural result of your affection for her . She used to be all unres  
distinguishing Elinor by no mark of affection . Marianne saw and listened with i  
th no inclination for expense , no affection for strangers , no profession , an  
till distinguishing her by the same affection which once she had felt no doubt o  
al of her confidence in Edward ' s affection , to the remembrance of every mark  
was made ? Had he never owned his affection to yourself ?" " Oh , no ; but if

---

-----  
-----  
Displaying 25 of 38 matches:

ay when they were created . And Adam lived an hundred and thirty years , and be  
ughters : And all the days that Adam lived were nine hundred and thirty yea and  
nd thirty yea and he died . And Seth lived an hundred and five years , and bega  
ve years , and begat Enos : And Seth lived after he begat Enos eight hundred an  
welue years : and he died . And Enos lived ninety years , and begat Cainan : An  
years , and begat Cainan : And Enos lived after he begat Cainan eight hundred  
ive years : and he died . And Cainan lived seventy years and begat Mahalaleel :  
rs and begat Mahalaleel : And Cainan lived after he begat Mahalaleel eight hund  
years : and he died . And Mahalaleel lived sixty and five years , and begat Jar  
s , and begat Jared : And Mahalaleel lived after he begat Jared eight hundred a  
and five yea and he died . And Jared lived an hundred sixty and two years , and  
o years , and he begat Eno And Jared lived after he begat Enoch eight hundred y  
and two yea and he died . And Enoch lived sixty and five years , and begat Met

; for God took him . And Methuselah lived an hundred eighty and seven years , , and begat Lamech . And Methuselah lived after he begat Lamech seven hundred nd nine yea and he died . And Lamech lived an hundred eighty and two years , an ch the LORD hath cursed . And Lamech lived after he begat Noah five hundred nin naan shall be his servant . And Noah lived after the flood three hundred and fi xad two years after the flo And Shem lived after he begat Arphaxad five hundred at sons and daughters . And Arphaxad lived five and thirty years , and begat Sa ars , and begat Salah : And Arphaxad lived after he begat Salah four hundred an begat sons and daughters . And Salah lived thirty years , and begat Eber : And y years , and begat Eber : And Salah lived after he begat Eber four hundred and begat sons and daughters . And Eber lived four and thirty years , and begat Pe y years , and begat Peleg : And Eber lived after he begat Peleg four hundred an

In [14]: #Trying Similar

```
text1.similar("monstrous")
text2.similar("monstrous")
```

true contemptible christian abundant few part mean careful puzzled  
mystifying passing curious loving wise doleful gamesome singular  
delightfully perilous fearless  
very so exceedingly heartily a as good great extremely remarkably  
sweet vast amazingly

In [ ]: #Trying common contexts

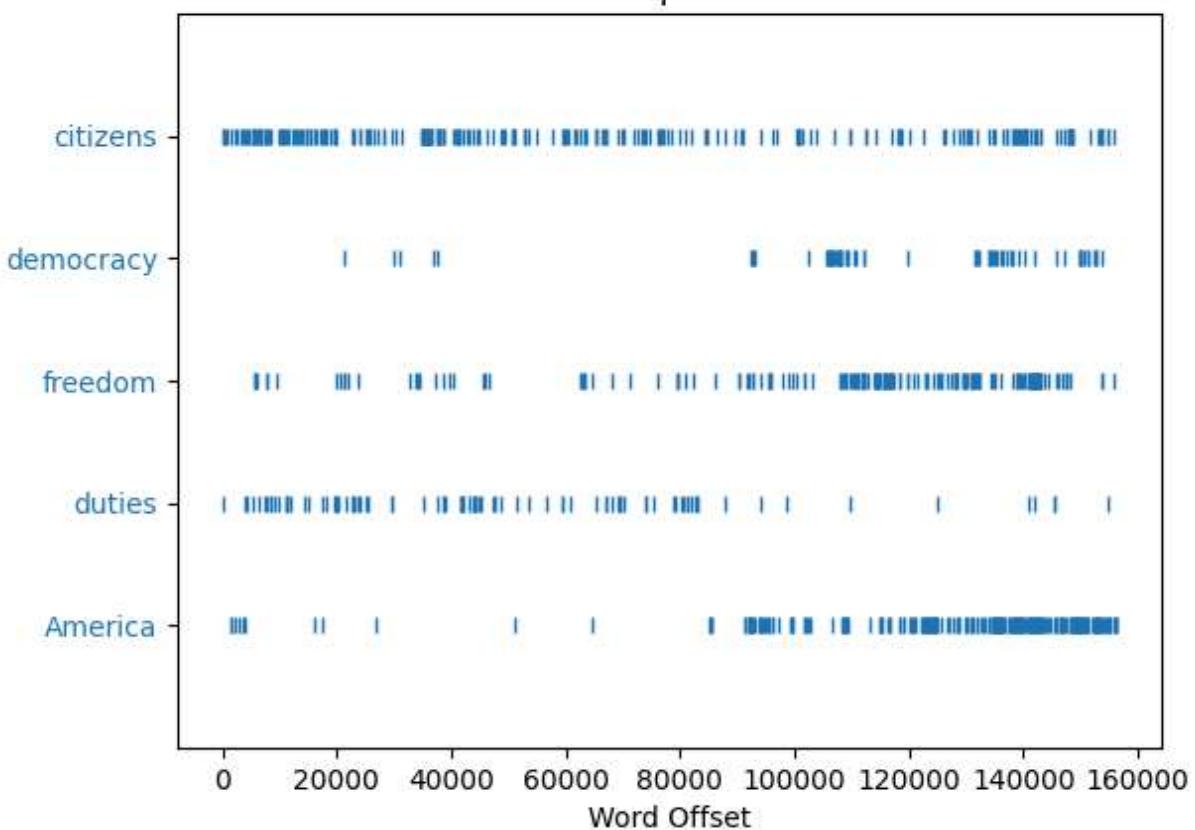
```
text2.common_contexts(["monstrous", "very"])
```

am\_glad a.pretty a\_lucky is\_pretty be\_glad

In [ ]: #Plots

```
text4.dispersion_plot(["citizens", "democracy", "freedom", "duties", "America"])
```

### Lexical Dispersion Plot



```
In [ ]: #Using Generate
text3.generate()
```

Building ngram index...

laid by her , and said unto Cain , Where art thou , and said , Go to ,  
I will not do it for ten ' s sons ; we dreamed each man according to  
their generatio the firstborn said unto Laban , Because I said , Nay ,  
but Sarah shall her name be . , duke Elah , duke Shobal , and Akan .  
and looked upon my affliction . Bashemath Ishmael ' s blood , but Isra  
for as a prince hast thou found of all the cattle in the valley , and  
the wo The

```
Out[ ]: "laid by her , and said unto Cain , Where art thou , and said , Go to ,\nI will no  
t do it for ten ' s sons ; we dreamed each man according to\ntheir generatio the f  
irstborn said unto Laban , Because I said , Nay ,\nbut Sarah shall her name be . ,  
duke Elah , duke Shobal , and Akan .\nand looked upon my affliction . Bashemath Is  
hmael ' s blood , but Isra\nfor as a prince hast thou found of all the cattle in t  
he valley , and\nthe wo The"
```

### QUESTION 2

```
In [25]: from nltk.tokenize import word_tokenize
nltk.download('punkt')
from collections import Counter
```

[nltk\_data] Downloading package punkt to  
[nltk\_data] C:\Users\DELL\AppData\Roaming\nltk\_data...  
[nltk\_data] Package punkt is already up-to-date!

```
In [ ]: #Assigning a string w paragraph
paragraph = """Natural language processing (NLP) is a subfield of linguistics, comp
```

```
In [ ]: #Counting Total and Unique Word counts
words = word_tokenize(paragraph.lower())
total_words = len(words)
unique_words = len(set(words))
```

```
In [ ]: #Word Frequencies
freq = Counter(words)
most_common = freq.most_common(1)
least_common = freq.most_common()[-1]
longest_word = max(words, key=len)
```

```
In [ ]: #Output
print("Total words:", total_words)
print("Unique words:", unique_words)
print("Most frequent word:", most_common)
print("Least frequent word:", least_common)
print("Longest word:", longest_word)
```

Total words: 28  
 Unique words: 25  
 Most frequent word: [('language', 2)]  
 Least frequent word: ('.', 1)  
 Longest word: intelligence

### QUESTION 3

```
In [27]: #Exercise 2.1
import re
re.findall(r'\b[a-zA-Z][a-zA-Z0-9_]*\b', 'abc _temp1 23start var_name')
```

Out[27]: ['abc', '\_temp1', 'var\_name']

```
In [ ]: #Exercise 2.2
text = "Don't hesitate to ask questions!"
re.findall(r"[A-Za-z]+(?:'[A-Za-z]+)?", text)
```

Out[ ]: ["Don't", 'hesitate', 'to', 'ask', 'questions']

### ADDITIONAL QUESTION

```
In [50]: #Practice Question 1
import string
import matplotlib.pyplot as plt
from nltk.corpus import stopwords

# Sample text
text = open("chapter.txt", encoding='utf-8').read().lower()
tokens = word_tokenize(text)
tokens = [word for word in tokens if word.isalpha()]

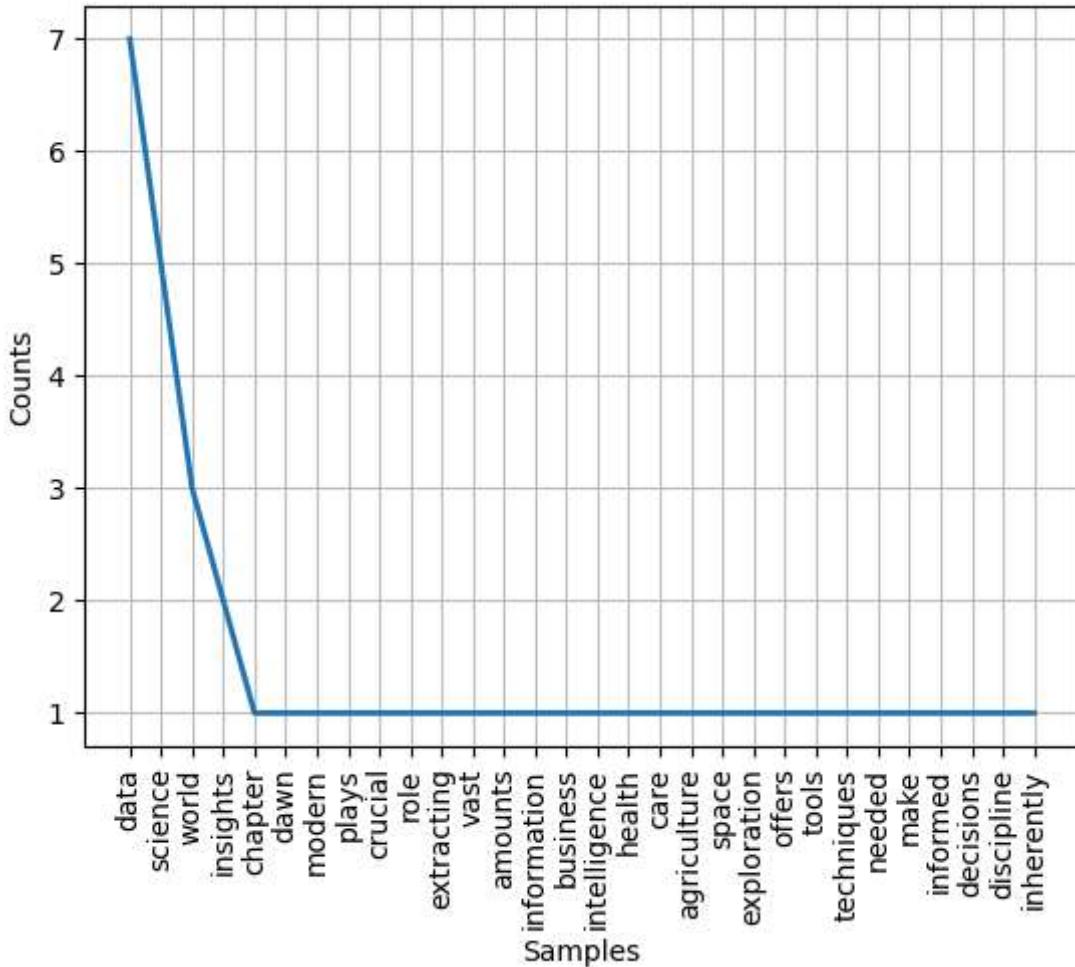
# Removing stopwords
```

```
nltk.download('stopwords')
stop_words = set(stopwords.words('english'))
filtered_tokens = [word for word in tokens if word not in stop_words]

# Frequency distribution
fdist = nltk.FreqDist(filtered_tokens)
fdist.plot(30, cumulative=False)
```

[nltk\_data] Downloading package stopwords to  
[nltk\_data] C:\Users\DELL\AppData\Roaming\nltk\_data...  
[nltk\_data] Package stopwords is already up-to-date!

Out[50]: <Axes: xlabel='Samples', ylabel='Counts'>



## Practice Question 2

In [41]: #Question 24  
# a. Ending in 'ize'  
words\_ize = [w for w in text6 if w.endswith('ize')]  
print("Words that ends with ize", words\_ize)

# b. Containing 'z'  
words\_z = [w for w in text6 if 'z' in w]  
print("Words that contain z", words\_z)

# c. Containing 'pt'

```
words_pt = [w for w in text6 if 'pt' in w]
print("Words containing pt",words_pt)

# d. Titlecase (initial capital)
words_title = [w for w in text6 if w.istitle() and w.isalpha()]
print("Words with Initial capital letters",words_title)
```

Words that ends with ize [ɪz]



'A', 'You', 'And', 'And', 'And', 'And', 'And', 'Yes', 'A', 'A', 'There', 'And', 'The', 'The', 'Well', 'I', 'Sir', 'Galahad', 'Oh', 'Quick', 'What', 'Quick', 'Why', 'You', 'No', 'Silence', 'You', 'Come', 'We', 'Look', 'I', 'Come', 'Sir', 'Galahad', 'No', 'Look', 'I', 'Yes', 'Let', 'Yes', 'Let', 'No', 'Sir', 'Galahad', 'Come', 'No', 'Really', 'Honestly', 'I', 'I', 'Oh', 'Let', 'Yes', 'Let', 'No', 'Quick', 'Quick', 'Please', 'I', 'There', 'Yes', 'He', 'We', 'He', 'Oh', 'We', 'You', 'I', 'I', 'Yes', 'You', 'Look', 'No', 'Look', 'I', 'No', 'Holy', 'Grail', 'Come', 'Oh', 'No', 'It', 'I', 'No', 'I', 'Sir', 'Launcelot', 'Sir', 'Galahad', 'Grail', 'Meanwhile', 'King', 'Arthur', 'Sir', 'Bedevere', 'Oh', 'I', 'I', 'Get', 'Oh', 'On', 'Arthur', 'I', 'Heh', 'Hee', 'And', 'Grail', 'Ha', 'Heh', 'Ha', 'Ha', 'Where', 'Heh', 'Old', 'Hee', 'He', 'And', 'Grail', 'The', 'Grail', 'There', 'Gorge', 'Eternal', 'Peril', 'But', 'Grail', 'Where', 'Grail', 'Seek', 'Bridge', 'Death', 'The', 'Bridge', 'Death', 'Grail', 'Heh', 'Ha', 'Hee', 'Ni', 'Ni', 'Ni', 'Ni', 'Ni', 'Who', 'We', 'Knights', 'Who', 'Say', 'Ni', 'Ni', 'No', 'Not', 'Knights', 'Who', 'Say', 'Ni', 'The', 'Who', 'We', 'Ni', 'Peng', 'Neee', 'Neee', 'Those', 'The', 'Knights', 'Who', 'Say', 'Ni', 'Knights', 'Ni', 'Ni', 'Ni', 'Ni', 'Ni', 'Ni', 'Ow', 'Ow', 'Ow', 'Agh', 'We', 'Well', 'We', 'A', 'Ni', 'Ni', 'Ni', 'Ow', 'Oh', 'Please', 'No', 'We', 'You', 'O', 'Knights', 'Ni', 'One', 'Of', 'And', 'Yes', 'Now', 'Hmm', 'Oh', 'Great', 'Hm', 'Hmm', 'Hm', 'Hmm', 'Ohh', 'Ay', 'Thsss', 'Ayy', 'Thsss', 'Ayy', 'Stop', 'Stop', 'Aay', 'Stop', 'Look', 'Clear', 'Go', 'Go', 'Go', 'And', 'Clear', 'Hah', 'Blood', 'The', 'Tale', 'Sir', 'Launcelot', 'One', 'What', 'No', 'Not', 'All', 'This', 'But', 'Mother', 'Father', 'Father', 'B', 'Father', 'I', 'Listen', 'I', 'When', 'I', 'Other', 'I', 'I', 'It', 'So', 'I', 'That', 'So', 'I', 'That', 'But', 'And', 'But', 'I', 'I', 'Rather', 'I', 'Stop', 'Stop', 'You', 'I', 'Now', 'In', 'Britain', 'B', 'I', 'Listen', 'Alice', 'Herbert', 'Erbert', 'We', 'We', 'But', 'I', 'Don', 'What', 'She', 'She', 'I', 'I', 'Cut', 'Cut', 'Look', 'Princess', 'Lucky', 'Guards', 'Make', 'Prince', 'I', 'Not', 'Hic', 'No', 'Until', 'I', 'Until', 'No', 'No', 'You', 'And', 'Hic', 'Right', 'We', 'No', 'Leaving', 'Leaving', 'Yes', 'All', 'Right', 'Hic', 'Right', 'Oh', 'Yes', 'What', 'Oh', 'Look', 'Uh', 'You', 'Alright', 'Hi', 'Right', 'Oh', 'I', 'Uhh', 'N', 'No', 'You', 'Oh', 'We', 'But', 'No', 'Just', 'Until', 'No', 'Just', 'Just', 'Hic', 'Get', 'Get', 'All', 'Right', 'We', 'Hic', 'And', 'What', 'Make', 'The', 'Prince', 'Yes', 'Make', 'Oh', 'Hic', 'Ah', 'I', 'You', 'Is', 'Hic', 'Oh', 'No', 'Right', 'Where', 'We', 'No', 'I', 'Oh', 'I', 'Right', 'But', 'Father', 'Shut', 'And', 'And', 'Hic', 'Oh', 'Well', 'Concorde', 'Thank', 'Most', 'And', 'Over', 'Good', 'Steady', 'And', 'Uuh', 'Come', 'Concorde', 'Message', 'Concorde', 'Concorde', 'Speak', 'To', 'I', 'Please', 'I', 'Tall', 'Tower', 'Swamp', 'Castle', 'At', 'A', 'A', 'This', 'Holy', 'Grail', 'Brave', 'Concorde', 'Uh', 'I', 'I', 'Well', 'I', 'I', 'I', 'I', 'Oh', 'I', 'Actually', 'I', 'I', 'No', 'Concorde', 'Stay', 'I', 'I', 'Idiom', 'Idiom', 'No', 'I', 'Farewell', 'Concorde', 'I', 'I', 'Shall', 'I', 'Yeah', 'Morning', 'Morning', 'Oooh', 'Ha', 'Hiyya', 'Hey', 'Hiyya', 'Ha', 'Ha', 'Huy', 'Uuh', 'Aaah', 'Ha', 'And', 'Aah', 'Hiyah', 'Aah', 'Aaah', 'Hyy', 'Hya', 'Hiyya', 'Ha', 'Now', 'O', 'Sir', 'Launcelot', 'Camelot', 'I', 'Oh', 'I', 'You', 'Uh', 'I', 'I', 'You', 'Uh', 'You', 'I', 'I', 'Well', 'I', 'Stop', 'Stop', 'Stop', 'Stop', 'Who', 'I', 'No', 'Uh', 'I', 'Sir', 'Launcelot', 'He', 'Father', 'Well', 'Did', 'Uh', 'Oh', 'Sorry', 'They', 'Well', 'I', 'Um', 'I', 'Don', 'Si', 'r', 'Launcelot', 'I', 'You', 'Well', 'I', 'I', 'Hurry', 'Sir', 'Launcelot', 'Hurry', 'Shut', 'You', 'Well', 'I', 'Didnt', 'You', 'Oh', 'Is', 'You', 'This', 'Well', 'I', 'I', 'Camelot', 'I', 'Camelot', 'Are', 'Camelot', 'Hurry', 'Sir', 'Launcelot', 'Uh', 'I', 'Knight', 'King', 'Arthur', 'Very', 'Camelot', 'Uh', 'Is', 'Hurry', 'I', 'Woul', 'd', 'Well', 'I', 'I', 'Um', 'Oooh', 'I', 'I', 'Oh', 'Oooh', 'Well', 'We', 'There', 'Oh', 'Ha', 'Hey', 'Ha', 'Hold', 'Stop', 'Hold', 'Hold', 'Hold', 'Hold', 'Hold', 'Please', 'Sorry', 'Sorry', 'You', 'I', 'I', 'Sorry', 'Sorry', 'He', 'Hold', 'Hold', 'Please', 'Hold', 'This', 'Sir', 'Launcelot', 'Court', 'Camelot', 'Hello', 'He', 'Please', 'Please', 'This', 'Let', 'We', 'Unfortunately', 'Herbert', 'Oh', 'Oh', 'But', 'I', 'I', 'For', 'He', 'Since', 'He', 'For', 'S', 'Uugh', 'Oh', 'And', 'I', 'And', 'I', 'Princess', 'Sir', 'Launcelot', 'Camelot', 'What', 'Look', 'The', 'Princ

e', 'Oooh', 'The', 'Prince', 'He', 'No', 'I', 'You', 'Tall', 'Tower', 'No', 'I', 'Ho w', 'Well', 'I', 'Not', 'No', 'Stop', 'He', 'He', 'Shut', 'He', 'Shut', 'He', 'Shut', 'He', 'Not', 'He', 'He', 'He', 'Quickly', 'He', 'Come', 'He', 'He', 'N o', 'It', 'He', 'I', 'Oh', 'Dramatically', 'Dramatically', 'But', 'Heeee', 'Hoa', 'Ho o', 'What', 'Excuse', 'Could', 'King', 'Arthur', 'Old', 'Is', 'Who', 'The', 'Knight s', 'Who', 'Say', 'Ni', 'Aggh', 'No', 'Never', 'We', 'If', 'I', 'Agh', 'Do', 'Very', 'If', 'No', 'Never', 'No', 'Ni', 'Nu', 'No', 'Nu', 'No', 'Nu', 'No', 'You', 'No', 'N i', 'Ni', 'That', 'That', 'You', 'Ni', 'Ohh', 'Ni', 'Agh', 'Ni', 'Ni', 'Ni', 'Ni', 'Ni', 'Are', 'Erm', 'Oh', 'There', 'Nothing', 'Even', 'Did', 'Yes', 'Shrubberi es', 'I', 'My', 'Roger', 'Shrubber', 'I', 'Ni', 'No', 'No', 'O', 'Knights', 'N i', 'May', 'It', 'I', 'But', 'What', 'We', 'Knights', 'Who', 'Say', 'Ni', 'Ni', 'Sh h', 'Shh', 'We', 'Knights', 'Who', 'Say', 'Ecky', 'Ni', 'Therefore', 'What', 'O', 'K nights', 'Knights', 'Who', 'Til', 'Recently', 'Said', 'Ni', 'Firstly', 'Not', 'Ni', 'Then', 'A', 'A', 'A', 'Ni', 'Shh', 'Ni', 'Ni', 'Shh', 'Then', 'We', 'O h', 'Cut', 'It', 'Aaaaugh', 'Aaaaugh', 'Augh', 'Ohh', 'Don', 'What', 'I', 'Knights', 'Ni', 'How', 'Aaaaugh', 'You', 'What', 'Agh', 'No', 'No', 'You', 'No', 'Not', 'My', 'Sir', 'Robin', 'Packing', 'And', 'And', 'Yes', 'Sir', 'Robin', 'My', 'It', 'Now', 'Surely', 'Holy', 'Grail', 'He', 'Shut', 'No', 'No', 'Far', 'He', 'Aaaaugh', 'I', 'A aaaaugh', 'Uh', 'No', 'Aaaaugh', 'Aaaaugh', 'Stop', 'The', 'Oh', 'Ow', 'He', 'Patsy', 'Wait', 'I', 'I', 'Ooh', 'I', 'And', 'That', 'Ohh', 'Aaaaugh', 'And', 'Arthur', 'Bed evere', 'Sir', 'Robin', 'Beyond', 'Launcelot', 'Galahad', 'Yay', 'Yay', 'In', 'Nador', 'Robin', 'Get', 'Eee', 'And', 'Yay', 'A', 'Winter', 'Spring', 'Mmm', 'Spring', 'Summer', 'Oh', 'Ahh', 'Summer', 'Winter', 'Oh', 'And', 'Winter', 'Spring', 'Summe r', 'Autumn', 'Aah', 'Oh', 'Waa', 'Until', 'King', 'Arthur', 'Eh', 'Oh', 'See', 'O h', 'Oh', 'Knights', 'Forward', 'What', 'I', 'By', 'There', 'Tim', 'Greetings', 'Ti m', 'Enchanter', 'Greetings', 'King', 'Arthur', 'You', 'I', 'You', 'Holy', 'Grail', 'That', 'You', 'O', 'Tim', 'Quite', 'Oh', 'Yes', 'Holy', 'Grail', 'Our', 'Holy', 'Grail', 'Yeah', 'Yes', 'It', 'It', 'Yeah', 'Yup', 'Yup', 'Hm', 'And', 'Yes', 'Yeah', 'We', 'We', 'We', 'Ages', 'Umhm', 'Uh', 'Look', 'Fine', 'Um', 'I', 'I', 'A', 'A', 'A', 'Yes', 'I', 'Y', 'Yes', 'Yup', 'That', 'Yes', 'Oh', 'Oh', 'Thank', 'Ahh', 'Oh', 'Fine', 'Thank', 'Splendid', 'Aah', 'Look', 'Yes', 'I', 'Holy', 'Grail', 'Oh', 'Oh', 'To', 'Caerbannog', 'Olfin', 'Bedwere', 'Rheged', 'Holy', 'Grail', 'Where', 'O', 'Tim', 'Follow', 'But', 'Follow', 'Bones', 'So', 'What', 'They', 'Then', 'Dis', 'Behol d', 'Caerbannog', 'Right', 'Keep', 'What', 'W', 'Too', 'What', 'There', 'Where', 'There', 'What', 'It', 'You', 'What', 'You', 'Well', 'Ohh', 'That', 'You', 'I', 'I', 'L ook', 'Get', 'He', 'Oh', 'You', 'I', 'What', 'He', 'Go', 'Bors', 'Chop', 'Right', 'S illy', 'One', 'Look', 'Aaaaugh', 'Jesus', 'Christ', 'I', 'I', 'I', 'Oh', 'Oh', 'Wel l', 'I', 'Oh', 'Do', 'Right', 'Oh', 'Charge', 'Aaaaugh', 'Aaaaugh', 'Run', 'Run', 'Ru n', 'Run', 'Ha', 'Ha', 'Ha', 'Ha', 'Right', 'How', 'Gawain', 'Ector', 'And', 'Bors', 'That', 'Three', 'Three', 'Three', 'And', 'That', 'Would', 'Oh', 'Let', 'It', 'Lik e', 'Well', 'Have', 'No', 'We', 'Holy', 'Hand', 'Grenade', 'Yes', 'The', 'Holy', 'Ha nd', 'Grenade', 'Antioch', 'Tis', 'Brother', 'Maynard', 'Brother', 'Maynard', 'Brin g', 'Holy', 'Hand', 'Grenade', 'Pie', 'Iesu', 'Pie', 'Iesu', 'Pie', 'Iesu', 'Pie', 'Iesu', 'How', 'I', 'Consult', 'Book', 'Armaments', 'Armaments', 'Chapter', 'Two', 'Nine', 'Twenty', 'And', 'Saint', 'Attila', 'O', 'Lord', 'And', 'Lord', 'Skip', 'Bro ther', 'And', 'Lord', 'First', 'Holy', 'Pin', 'Then', 'Three', 'Four', 'Five', 'One', 'Onc e', 'Holy', 'Hand', 'Grenade', 'Antioch', 'Amen', 'Amen', 'Right', 'One', 'Three', 'Three', 'There', 'Look', 'What', 'What', 'Brother', 'Maynard', 'You', 'It', 'Aramai c', 'Of', 'Joseph', 'Arimathea', 'Course', 'What', 'It', 'Here', 'Joseph', 'Arimathea', 'He', 'Holy', 'Grail', 'Castle', 'What', 'Castle', 'What', 'He', 'Oh', 'Well', 'Look', 'He', 'Well', 'Perhaps', 'Oh', 'Well', 'No', 'Just', 'Aauuggghhh', 'Aaaugg h', 'Do', 'Camaaaaaargue', 'Where', 'France', 'I', 'Isn', 'Saint', 'Auuuuvves', 'Corn wall', 'No', 'Saint', 'Ives', 'Oh', 'Saint', 'Iiiives', 'Iiiives', 'Oooohooohoooo', 'No', 'Auuuuuugh', 'Auuuugh', 'N', 'No', 'Oooooooh', 'Oh', 'Yes', 'I', 'Ooh', 'My', 'God', 'It', 'Black', 'Beast', 'Aaaaugh', 'Black', 'Beast', 'Aaaaugh', 'That', 'Tha t', 'Run', 'Run', 'Run', 'Run', 'Keep', 'Shh', 'Shh', 'Shh', 'Shh', 'Shh', 'S

```
hh', 'Shh', 'Shh', 'We', 'Aagh', 'As', 'Black', 'Beast', 'Arthur', 'Ulk', 'The', 'The', 'Holy', 'Grail', 'There', 'The', 'Bridge', 'Death', 'Oh', 'Look', 'There', 'What', 'He', 'Bridge', 'Death', 'He', 'Three', 'Three', 'He', 'Three', 'Three', 'Three', 'What', 'Then', 'Gorge', 'Eternal', 'Peril', 'Oh', 'I', 'Who', 'Sir', 'Robin', 'Yes', 'Brave', 'Sir', 'Robin', 'Hey', 'I', 'Why', 'Launcelot', 'Yes', 'Let', 'I', 'I', 'No', 'No', 'Hang', 'Hang', 'Hang', 'Just', 'Three', 'Three', 'And', 'I', 'Good', 'Sir', 'Launcelot', 'God', 'Stop', 'Who', 'Bridge', 'Death', 'Ask', 'I', 'What', 'My', 'Sir', 'Launcelot', 'Camelot', 'What', 'To', 'Holy', 'Grail', 'What', 'Blue', 'Right', 'Off', 'Oh', 'Thank', 'That', 'Stop', 'Who', 'Bridge', 'Death', 'Ask', 'I', 'What', 'Sir', 'Robin', 'Camelot', 'What', 'To', 'Holy', 'Grail', 'What', 'Assyria', 'I', 'Auuuuuugh', 'Stop', 'What', 'Sir', 'Galahad', 'Camelot', 'What', 'I', 'Grail', 'What', 'Blue', 'No', 'Hee', 'Stop', 'What', 'It', 'Arthur', 'King', 'Britons', 'What', 'To', 'Holy', 'Grail', 'What', 'What', 'An', 'African', 'European', 'Huh', 'I', 'I', 'Auuuuuuuugh', 'How', 'Well', 'Launcelot', 'Launcelot', 'Launcelot', 'Launcelot', 'Launcelot', 'Launcelot', 'Launcelot', 'Launcelot', 'The', 'Castle', 'Aaaagh', 'Our', 'God', 'Almighty', 'God', 'Thee', 'Thou', 'Jesus', 'Christ', 'Allo', 'English', 'Monsieur', 'Arthur', 'King', 'So', 'French', 'How', 'I', 'Knights', 'Camelot', 'God', 'Himself', 'How', 'English', 'I', 'So', 'French', 'I', 'In', 'Lord', 'No', 'English', 'I', 'If', 'In', 'God', 'Agh', 'Right', 'That', 'Yes', 'Ha', 'Walk', 'Just', 'And', 'And', 'English', 'Thpppt', 'We', 'Yes', 'Stand', 'French', 'Dappy', 'Today', 'In', 'God', 'Hoo', 'Ohh', 'Holy', 'Grail', 'God', 'Ha', 'Charge', 'Hoory', 'S', 'Yes', 'They', 'I', 'Come', 'Anybody', 'All', 'Come', 'Back', 'S', 'Get', 'Back', 'Right', 'Just', 'Come', 'Come', 'Put', 'Clear', 'Come', 'With', 'Which', 'Oh', 'Come', 'Put', 'Get', 'We', 'Ahh', 'Ooh', 'Come', 'Back', 'Riiight', 'Come', 'Run', 'Run', 'Pull', 'My', 'Come', 'Back', 'Back', 'Right', 'Come', 'Everything', 'All', 'That', 'Just', 'Christ']
```

```
In [ ]: #Question 25
sent = ['she', 'sells', 'sea', 'shells', 'by', 'the', 'sea', 'shore']

# a. Words starting with 'sh'
sh_words = [w for w in sent if w.startswith('sh')]

print("Words that begins with sh", sh_words)

# b. Words Longer than 4 characters
long_words = [w for w in sent if len(w) > 4]
print("Words which are longer than 4 characters", long_words)
```

Words that begins with sh ['she', 'shells', 'shore']  
 Words which are longer than 4 characters ['sells', 'shells', 'shore']

```
In [ ]: #Question 26
total_chars = sum(len(w) for w in text1)
total_words = len(text1)
average_word_length = total_chars / total_words

print(f"Average word length: {average_word_length:.2f}")

#Iterates over all words w in text1
#for each word w, it calculates len(w) i.e. the number of characters in the word.
#sum() adds up these lengths
```

Average word length: 3.83

```
In [ ]: #Question 27
def vocab_size(text):
    return len(set(text))
```

```
print("Vocabulary size of text1 (Moby Dick):", vocab_size(text1))
```

Vocabulary size of text1 (Moby Dick): 19317

```
In [48]: #Question 28
def percent(word, text):
    return 100 * text.count(word) / len(text)

print("Percentage of 'whale' in text1:", percent("whale", text1))
```

Percentage of 'whale' in text1: 0.3473673313677301

In [ ]: