NLTK

AIM:
Import NLTK
Import any text from the book
Perform The following tasks:

1. Concordance
2. Common_context
3. Generate
4. Count
5. Similar
6. Dispertion_plot
7. Implement n gram model ----2nd lab
8. Tfidf implementation-3rd lab

In [2]:

```python
import nltk
nltk.download()
```

showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml (https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml)

Out[2]:

True

In [3]:

```python
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
```

In [4]:

```
text1
```

Out[4]:

```
<Text: Moby Dick by Herman Melville 1851>
```

In [5]:

```python
from nltk.corpus import gutenberg
from nltk.text import Text
```

In [6]:

```python
corpus = gutenberg.words('melville-moby_dick.txt')
```

In [7]:

```python
text = Text(corpus)
```

Concordance

In [8]:

```python
text.concordance("monstrous")
```

```
Displaying 11 of 11 matches:
ong the former , one was of a most monstrous size . ... This came toward
s 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 wer
e thick
d as you gazed , and wondered what monstrous cannibal and savage could e
ver hav
that has survived the flood ; most monstrous and most mountainous ! That
Himmal
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 shal
l ere l
ing Scenes . In connexion with the monstrous pictures of whales , I am s
trongly
ere to enter upon those still more monstrous stories of them which are t
o 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
```

In [9]:

```python
from nltk.tokenize.regexp import WhitespaceTokenizer
```

In [10]:

```python
tokens = WhitespaceTokenizer().tokenize(str(text))
```

In [11]:

```python
print(len(tokens))
```

7

In [12]:

```python
text.concordance("Moby")
```

```
Displaying 25 of 84 matches:
[ Moby Dick by Herman Melville 1851 ] ETYMO
hale must be the same that some call Moby Dick ." " Moby Dick ?" sh
outed Ahab
e same that some call Moby Dick ." " Moby Dick ?" shouted Ahab . "
Do ye know
all . Death and devils ! men , it is Moby Dick ye have seen -- Moby
Dick -- Mo
en , it is Moby Dick ye have seen -- Moby Dick -- Moby Dick !" " Ca
ptain Ahab
by Dick ye have seen -- Moby Dick -- Moby Dick !" " Captain Ahab ,"
said Starb
r . " Captain Ahab , I have heard of Moby Dick -- but it was not Mo
by Dick tha
heard of Moby Dick -- but it was not Moby Dick that took off thy le
g ?" " Who
aye , my hearties all round ; it was Moby Dick that dismasted me ;
Moby Dick t
it was Moby Dick that dismasted me ; Moby Dick that brought me to t
```

Common Text

In [23]:

```python
from nltk.book import text7
# define the target word
target_word = "Washington"

# find the common contexts for the target word
common_contexts = text7.common_contexts([target_word])
common_contexts
```
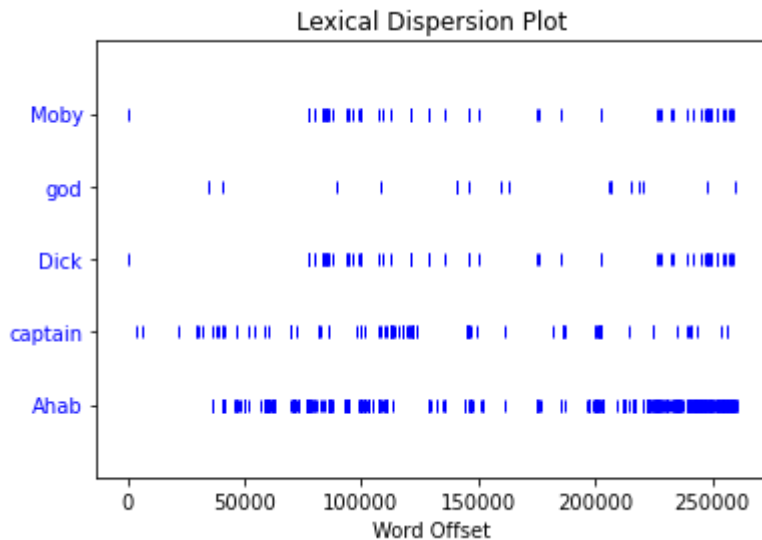
```
in_, in_and the_post a_lawyer mary_college reasserts_'s about_'s the_,
as_; in_have in_. that_and in_that and_. in_for to_every and_, to_,
```

Dispersion Plot

In [34]:

```python
# define the words to plot
words_to_plot = ['Moby', 'god', 'Dick', 'captain', 'Ahab']
# generate the dispersion plot
text1.dispersion_plot(words_to_plot)
```


Generate

In [25]:

```python
from nltk.parse.generate import generate, demo_grammar
```

In [54]:

```python
text = nltk.Text(text1)
# generate 20 words of random text
random_text = text.generate(25)
random_text
```

Building ngram index...

long , from one to the top - mast , and no coffin and went out a sea
captain -- this peaking of the whales

Out[54]:

'long , from one to the top - mast , and no coffin and went out a sea\nc
aptain -- this peaking of the whales'

In [28]:

```python
random_text_sample = ['the', 'quick', 'brown', 'fox', 'jumps']
sentence = ' '.join(random_text_sample) # to join the words to form a sentence
print(sentence)
```

the quick brown fox jumps

Count

In [40]:

```python
text = nltk.Text(text1)
count = text.count('an')
print(count)
```

582

Similar

In [41]:

```python
similar_words = text.similar('monstrous')
print(similar_words)
```

true contemptible christian abundant few part mean careful puzzled
mystifying passing curious loving wise doleful gamesome singular
delightfully perilous fearless
None

In [45]:

```python
similar_words = text7.similar('Street')
print(similar_words)
```

market company u.s. industry stock program fields year information
president securities government senate country economy case spinoff
development firm law
None

N-Gram Model

In [48]:

```python
from nltk.util import ngrams
from nltk.book import text6
text6 = nltk.Text(text6)
```

In [56]:

```python
# define the value of n for the N-gram model
n = int(input('Enter The Value of N'))
# create a list of all the n-grams in the text
n_grams = list(ngrams(text, n))
# count the frequency of each n-gram in the text
freq_dist = nltk.FreqDist(n_grams)
# print the 20 most common n-grams in the text
print(freq_dist.most_common(20))
```

```
Enter The Value of N5
[((',"', 'said', 'I', ',', '"'), 20), (('of', 'the', 'whale', "'", 's'),
17), (('the', 'Sperm', 'Whale', "'", 's'), 17), (('the', 'old', 'man',
"'", 's'), 15), ((',', 'as', 'it', 'were', ','), 13), (('I', 'don', "'",
't', 'know'), 11), (('the', 'bottom', 'of', 'the', 'sea'), 9), (('Queequ
eg', ',"', 'said', 'I', ','), 9), (('"', 'Aye', ',', 'aye', ','), 9),
((',', 'd', "'", 'ye', 'see'), 9), (('"', 'I', 'don', "'", 't'), 9),
(('the', 'sperm', 'whale', "'", 's'), 9), (('of', 'the', 'Pequod', "'",
's'), 9), (('and', 'at', 'the', 'same', 'time'), 9), (('of', 'the', 'Spe
rm', 'Whale', ','), 9), ((',', 'A', '.', 'D', '.'), 8), (('the', 'quarte
r', '-', 'deck', ','), 8), ((',', 'and', 'at', 'the', 'same'), 8), (('th
e', 'White', 'Whale', "'", 's'), 8), (('in', 'the', 'boat', "'", 's'),
8)]
```