# exp2

October 31, 2021

Neha Kulkarni

A060 BTech. IT SEM 7

Computational Linguistics and Natural Language Processing Lab Experiment 2

```
[1]: import nltk
```

```
[2]: 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
```

```
[3]: text1
```

```
[3]: <Text: Moby Dick by Herman Melville 1851>
```

```
[4]: text1.concordance('monstrous')
```

```
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 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 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
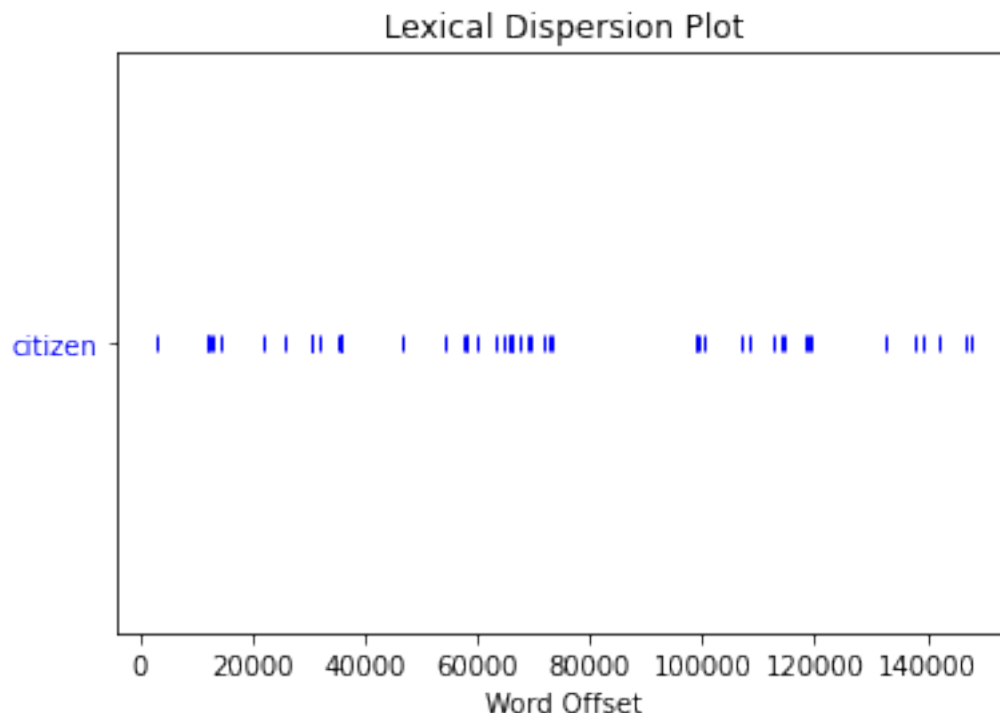of Whale - Bones ; for Whales of a monstrous size are oftentimes cast up dead u

[5]: `text1.similar("wall")`

whale sea world boat water ship air time wind forecastle door pequod
leviathan deep night line crew deck sun cabin

[6]: `text1.common_contexts(['monstrous','very'])`

No common contexts were found

[7]: `text4.dispersion_plot(['citizen'])`



Lexical Dispersion Plot

[8]: `text4.generate()`

Building ngram index…

occur , and especially the truth that democratic government has innate
capacity to govern its affairs aright through the Province ceded , by
any timid forebodings of evil were not to overtake them while I
possess the property of the peaks and the exercise of free and firm on
the farm , in view that the best of my countrymen will ever find me
ready to confer their benefits on countless generations yet to make
its promise for all generations ." , remains essentially unchanged .

```
cost of the Rocky Mountains . abuses of an ever - expanding American
dream
```

[8]: 'occur , and especially the truth that democratic government has
innate\ncapacity to govern its affairs aright through the Province ceded ,
by\nany timid forebodings of evil were not to overtake them while I\npossess the
property of the peaks and the exercise of free and firm on\nthe farm , in view
that the best of my countrymen will ever find me\nready to confer their benefits
on countless generations yet to make\nits promise for all generations ." ,
remains essentially unchanged .\ncost of the Rocky Mountains . abuses of an ever
- expanding American\ndream'

[9]: ```
len(text5)
```

[9]: 45010

[10]: ```
len(set(text5))
```

[10]: 6066

[11]: ```
len(set(text3)) / len(text3)
```

[11]: 0.06230453042623537

[12]: ```
100 * text4.count('a') / len(text4)
```

[12]: 1.457973123627309

[13]: ```
100*text5.count('lol')/len(text5)
```

[13]: 1.5640968673628082

[14]: ```
def lexical_diversity(text):
    return len(set(text)) / len(text)
def percentage(count, total):
    return 100 * count / total
```

[15]: ```
print(lexical_diversity(text3),percentage(5,6),percentage(text4.
 ↪count('q'),len(text4)))
```

0.06230453042623537 83.33333333333333 0.0

[ ]: