

NLP 2 Gayathri

```
In [1]:  text1="In a quaint town nestled between rolling hills and babbling brook
import nltk
fd=nltk.FreqDist(text1.split())
```

```
In [2]:  fd
```

```
Out[2]:  FreqDist({'the': 7, 'and': 5, 'a': 4, 'of': 4, 'through': 2, 'warm': 2, 'with': 2, 'horizon': 2, 'in': 2, 'where': 2, ...})
```

```
In [3]:  from nltk.probability import ConditionalFreqDist
cfd=ConditionalFreqDist((len(word),word)for word in text1.split())
```

```
In [5]:  cfd[5]
```

```
Out[5]:  FreqDist({'where': 2, 'hills': 1, 'pace.': 1, 'local': 1, 'array': 1, 'below': 1, 'comes': 1, 'alive': 1, 'soft': 1, 'spent': 1, ...})
```

```
In [6]:  pip install jieba
```

```
Collecting jieba
  Using cached jieba-0.42.1-py3-none-any.whl
Installing collected packages: jieba
Successfully installed jieba-0.42.1
Note: you may need to restart the kernel to use updated packages.
```

```
In [7]:  import jieba
```

```
In [8]:  seg_list=jieba.cut("水果店员工",cut_all=True)
```

```
In [9]:  print(seg_list)
```

```
<generator object Tokenizer.cut at 0x000002AE0817C5C0>
```

```
In [10]: print(",".join(seg_list))
```

```
Building prefix dict from the default dictionary ...
Dumping model to file cache C:\Users\gayus\AppData\Local\Temp\jieba.cache
Loading model cost 1.141 seconds.
Prefix dict has been built successfully.

水果,水果店,店员,员工
```

```
In [ ]:  
```

