In [2]:
```
text1='the baisi for the work is melvilles'
import nltk
fd=nltk.FreqDist(text1.split())
```

In [3]:
```
fd
```

Out[3]:
```
FreqDist({'the': 2, 'baisi': 1, 'for': 1, 'work': 1, 'is': 1, 'melvilles':
1})
```

In [4]:
```
from nltk.corpus import inaugural
inaugural.fileids()
```

Out[4]:
```
['1789-Washington.txt',
 '1793-Washington.txt',
 '1797-Adams.txt',
 '1801-Jefferson.txt',
 '1805-Jefferson.txt',
 '1809-Madison.txt',
 '1813-Madison.txt',
 '1817-Monroe.txt',
 '1821-Monroe.txt',
 '1825-Adams.txt',
 '1829-Jackson.txt',
 '1833-Jackson.txt',
 '1837-VanBuren.txt',
 '1841-Harrison.txt',
 '1845-Polk.txt',
 '1849-Taylor.txt',
 '1853-Pierce.txt',
 '1857-Buchanan.txt',
 '1861-Lincoln.txt',
 '1865-Lincoln.txt',
 '1869-Grant.txt',
 '1873-Grant.txt',
 '1877-Hayes.txt',
 '1881-Garfield.txt',
 '1885-Cleveland.txt',
 '1889-Harrison.txt',
 '1893-Cleveland.txt',
 '1897-McKinley.txt',
 '1901-McKinley.txt',
 '1905-Roosevelt.txt',
 '1909-Taft.txt',
 '1913-Wilson.txt',
 '1917-Wilson.txt',
 '1921-Harding.txt',
 '1925-Coolidge.txt',
 '1929-Hoover.txt',
 '1933-Roosevelt.txt',
 '1937-Roosevelt.txt',
 '1941-Roosevelt.txt',
 '1945-Roosevelt.txt',
 '1949-Truman.txt',
 '1953-Eisenhower.txt',
 '1957-Eisenhower.txt',
 '1961-Kennedy.txt',
 '1965-Johnson.txt',
 '1969-Nixon.txt',
 '1973-Nixon.txt',
 '1977-Carter.txt',
 '1981-Reagan.txt',
 '1985-Reagan.txt',
 '1989-Bush.txt',
 '1993-Clinton.txt',
 '1997-Clinton.txt',
 '2001-Bush.txt',
 '2005-Bush.txt',
 '2009-Obama.txt',
 '2013-Obama.txt',
```

```
 '2017-Trump.txt',
 '2021-Biden.txt']
```

In [5]:
```
text1='2021-Biden.txt'
import nltk
fd=nltk.FreqDist(text1.split())
```

In [6]:
```
fd
```

Out[6]:
```
FreqDist({'2021-Biden.txt': 1})
```

In [11]:
```
text1='hey hoe are you'
from nltk.probability import ConditionalFreqDist
cfd=ConditionalFreqDist((len(word),word)for word in text1.split())
```

In [12]:
```
cfd
```

Out[12]:
```
<ConditionalFreqDist with 1 conditions>
```

In [15]:
```
pip install jieba
```

```
8:04
     ---------------------------------------- 0.0/19.2 MB 6.7 kB/s eta 0:4
8:04
     ---------------------------------------- 0.0/19.2 MB 6.7 kB/s eta 0:4
8:04
     ---------------------------------------- 0.0/19.2 MB 6.7 kB/s eta 0:4
8:04
     ---------------------------------------- 0.0/19.2 MB 6.7 kB/s eta 0:4
8:04
     ---------------------------------------- 0.0/19.2 MB 6.7 kB/s eta 0:4
8:04
     ---------------------------------------- 0.0/19.2 MB 6.7 kB/s eta 0:4
8:04
     ---------------------------------------- 0.0/19.2 MB 6.7 kB/s eta 0:4
8:04
     ---------------------------------------- 0.0/19.2 MB 6.7 kB/s eta 0:4
8:04
     ---------------------------------------- 0.0/19.2 MB 6.7 kB/s eta 0:4
8:04
     ---------------------------------------- 0.0/19.2 MB 6.7 kB/s eta 0:4
```

In [16]:
```
import jieba
```

In [17]:
```
seg_list = jieba.cut("自然语言", cut_all=True)
```

In [18]:
```
print(seg_list)
```

```
<generator object Tokenizer.cut at 0x0000021FED76D900>
```

In [19]:
```
print(", ".join(seg_list))
```

```
Building prefix dict from the default dictionary ...
Dumping model to file cache C:\Users\SAYEE\AppData\Local\Temp\jieba.cache
Loading model cost 1.992 seconds.
Prefix dict has been built successfully.
```

自然，自然语言，语言

In [ ]:
```

```