

```
In [1]: text1="The All India Services (AIS) comprises three civil services in India
import nltk
fd=nltk.FreqDist(text1.split())
```

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

```
Out[2]: FreqDist({'the': 6, 'India': 3, 'Indian': 3, 'Service': 3, 'All': 2, 'Serv
ices': 2, 'civil': 2, 'in': 2, 'to': 2, 'centre': 2, ...})
```

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

```
In [4]: cfd[14]
```

```
Out[4]: FreqDist({'Administrative': 1})
```

```
In [5]: pip install jieba
```

```
0:00:11 ----- 12.8/19.2 MB 630.1 kB/s eta
0:00:11 ----- 12.8/19.2 MB 628.9 kB/s eta
0:00:11 ----- 12.9/19.2 MB 628.9 kB/s eta
0:00:11 ----- 12.9/19.2 MB 627.7 kB/s eta
0:00:11 ----- 12.9/19.2 MB 627.1 kB/s eta
0:00:11 ----- 12.9/19.2 MB 626.5 kB/s eta
0:00:11 ----- 12.9/19.2 MB 625.9 kB/s eta
0:00:11 ----- 13.0/19.2 MB 625.3 kB/s eta
0:00:10 ----- 13.0/19.2 MB 624.7 kB/s eta
0:00:10 ----- 13.0/19.2 MB 623.5 kB/s eta
```

```
In [ ]:
```

```
In [6]: import jieba
```

```
In [7]: seg_list=jieba.cut("自然语言",cut_all=True)
```

```
In [8]: print(seg_list)
```

```
<generator object Tokenizer.cut at 0x000001729A908720>
```

```
In [9]: print(",".join(seg_list))
```

```
Building prefix dict from the default dictionary ...  
Dumping model to file cache C:\Users\Vidhisha\AppData\Local\Temp\jieba.cache  
he  
Loading model cost 0.918 seconds.  
Prefix dict has been built successfully.
```

自然,自然语言,语言

```
In [ ]: PP
```