练习答案-第 7 章 _ 文本分析处理

2023 年 9 月 18 日

练习答案

练习：关于美国总统就职演说中提及“战争”的次数

1. 请问 Biden 的就职演说中关于关键词’war’ 的频数有多少?
2. 请绘制历届总统的就职演说中关键词’war’ 的频数随年份的变化。
   Tips: 注意，每个文本的年份都出现在其文件名中。使用 fileid[:4]，提取前四个字符，即可获得年份。

[2]:
```
from nltk.tokenize import word_tokenize
from nltk.corpus import inaugural
sentence = ' '.join(inaugural.words('2021-Biden.txt'))
sentence
fdist = {}
for word in word_tokenize(sentence):
    w = word.lower()
    if w not in fdist: fdist[w] = 0
    fdist[w] += 1
```

[3]:
```
fdist['war']
```

[3]: 7

[14]:
```
from nltk.tokenize import word_tokenize
from nltk.corpus import inaugural
fdist = {}
for i in inaugural.fileids():
    sentence = ' '.join(inaugural.words(i))
    fdist[i[:4]] = 0
```

```
    for word in word_tokenize(sentence):
        w = word.lower()
        if w=='war':
            fdist[i[:4]] += 1
```

[46]:

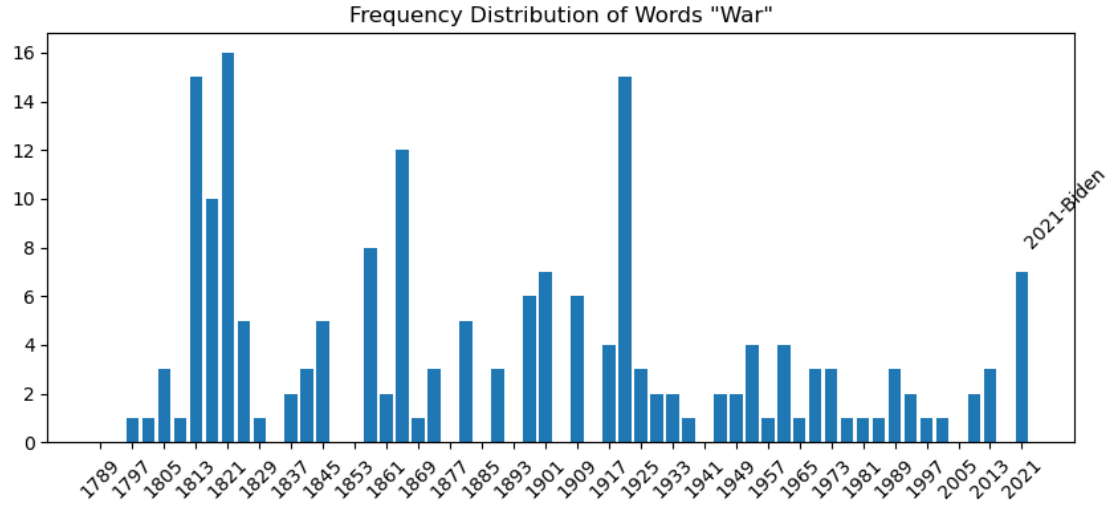
```
array(['1789', '1793', '1797', '1801', '1805', '1809', '1813', '1817',
       '1821', '1825', '1829', '1833', '1837', '1841', '1845', '1849',
       '1853', '1857', '1861', '1865', '1869', '1873', '1877', '1881',
       '1885', '1889', '1893', '1897', '1901', '1905', '1909', '1913',
       '1917', '1921', '1925', '1929', '1933', '1937', '1941', '1945',
       '1949', '1953', '1957', '1961', '1965', '1969', '1973', '1977',
       '1981', '1985', '1989', '1993', '1997', '2001', '2005', '2009',
       '2013', '2017', '2021'], dtype='<U4')
```

[67]:

```
import matplotlib.pyplot as plt
import numpy as np
fig, ax= plt.subplots(figsize=(10,4))
ax.bar(fdist.keys(), fdist.values())
n_sample = len(fdist)
xticks = range(0, n_sample, 2)
ax.set_xticks(xticks, np.array(list(fdist.keys()))[xticks], rotation = 45)
ax.set_title('Frequency Distribution of Words "War"')
ax.text('2021', 8, '2021-Biden', rotation=45)  # 添加文本
```

[67]:

```
Text(2021, 8, '2021-Biden')
```


Frequency Distribution of Words "War"
16
14
12
10
8
6
4
2
0
1789
1797
1805
1813
1821
1829
1837
1845
1853
1861
1869
1877
1885
1893
1901
1909
1917
1925
1933
1941
1949
1957
1965
1973
1981
1989
1997
2005
2013
2021
2021-Biden