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

In [2]:
```
nltk.download()
```

```
showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml (https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml)
```

Out[2]:
```
True
```

In [3]:
```
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
```

In [7]:
```
# Task 2: Imort Brown Corpus and Accessing Data
from nltk.corpus import brown

# categories in brown corpora
#brown.categories()
#brown.words(categories='adventure')[:20]
brown.words(categories='fiction')[:20]
```

Out[7]:
```
['Thirty-three',
 'Scotty',
 'did',
 'not',
 'go',
 'back',
 'to',
 'school',
 '.',
 'His',
 'parents',
 'talked',
 'seriously',
 'and',
 'lengthily',
 'to',
 'their',
 'own',
 'doctor',
 'and']
```

In [11]:
```
# Task 3: Import Inaugral Corpus and access data
# includes every president's inaugral address from 1789 to 2009
from nltk.corpus import inaugural
inaugural.fileids()
```

Out[11]:
```
['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']
```

In [15]:
```
inaugural.words(fileids = '2017-Trump.txt')[:50]
```

Out[15]:
```
['Chief',
 'Justice',
 'Roberts',
 ',',
 'President',
 'Carter',
 ',',
 'President',
 'Clinton',
 ',',
 'President',
 'Bush',
 ',',
 'President',
 'Obama',
 ',',
 'fellow',
 'Americans',
 ',',
 'and',
 'people',
 'of',
 'the',
 'world',
 ':',
 'Thank',
 'you',
 '.',
 'We',
 ',',
 'the',
 'citizens',
 'of',
 'America',
 ',',
 'are',
 'now',
 'joined',
 'in',
 'a',
 'great',
 'national',
 'effort',
 'to',
 'rebuild',
 'our',
 'country',
 'and',
 'restore',
 'its']
```

In [14]:
```
inaugural.words(fileids = '1861-Lincoln.txt')[:50]
```

Out[14]:
```
['Fellow',
 '-',
 'Citizens',
 'of',
 'the',
 'United',
 'States',
 ':',
 'In',
 'compliance',
 'with',
 'a',
 'custom',
 'as',
 'old',
 'as',
 'the',
 'Government',
 'itself',
 ',',
 'I',
 'appear',
 'before',
 'you',
 'to',
 'address',
 'you',
 'briefly',
 'and',
 'to',
 'take',
 'in',
 'your',
 'presence',
 'the',
 'oath',
 'prescribed',
 'by',
 'the',
 'Constitution',
 'of',
 'the',
 'United',
 'States',
 'to',
 'be',
 'taken',
 'by',
 'the',
 'President']
```

In [16]:
```
inaugural.words(fileids = '2009-Obama.txt')[:50]
print(inaugural.words(fileids = '2009-Obama.txt'))
```

Out[16]:
```
['My',
 'fellow',
 'citizens',
 ':',
 'I',
 'stand',
 'here',
 'today',
 'humbled',
 'by',
 'the',
 'task',
 'before',
 'us',
 ',',
 'grateful',
 'for',
 'the',
 'trust',
 'you',
 'have',
 'bestowed',
 ',',
 'mindful',
 'of',
 'the',
 'sacrifices',
 'borne',
 'by',
 'our',
 'ancestors',
 '.',
 'I',
 'thank',
 'President',
 'Bush',
 'for',
 'his',
 'service',
 'to',
 'our',
 'nation',
 ',',
 'as',
 'well',
 'as',
 'the',
 'generosity',
 'and',
 'cooperation']
```

In [17]:
```
#Task 4: Importing WEBTEXT CORPUS and Access Data
from nltk.corpus import webtext
webtext.fileids()

for fileid in webtext.fileids():
    print(fileid, webtext.raw(fileid)[:])
```

```
firefox.txt Cookie Manager: "Don't allow sites that set removed cookies to se
t future cookies" should stay checked
When in full screen mode
Pressing Ctrl-N should open a new browser when only download dialog is left o
pen
add icons to context menu
So called "tab bar" should be made a proper toolbar or given the ability coll
apse / expand.
[XUL] Implement Cocoa-style toolbar customization.
#ifdefs for MOZ_PHOENIX
customize dialog's toolbar has small icons when small icons is not checked
nightly builds and tinderboxen for Phoenix
finish tearing prefs UI to pieces and then make it not suck
"mozbrowser" script doesn't start correct binary
Need bookmark groups icon
Dropping at top of palette box horks things
keyboard shortcut for Increase Text Size is broken
default phoenix bookmarks
[cust] need a toolbar spacer and spring spacer for customize
```

In [18]:
```
# Task 5: Frequency Distribution of words in a text
text1 = '''1962 Tour de France was the 49th edition of the Tour de France, one of
fd = nltk.FreqDist(text1.split())
```

In [21]:
```
fd
```

Out[21]:
```
FreqDist({'the': 6, 'of': 5, 'Tour': 4, 'de': 3, 'was': 3, 'in': 3, 'and': 3,
'stages,': 2, 'on': 2, 'his': 2, ...})
```

In [23]:
```
# Task 6. Conditional Frequency Distribution of words in a text
# tells us how many 2 letter words or 3 letter words
from nltk.probability import ConditionalFreqDist
cfd = ConditionalFreqDist((len(word), word) for word in text1.split())
cfd[3]
```

Out[23]:
```
FreqDist({'the': 6, 'was': 3, 'and': 3, 'his': 2, 'one': 1, 'The': 1, 'mi)': 1,
'two': 1, 'des': 1, 'won': 1, ...})
```

In [24]:
```
cfd[6]
```

Out[24]:
```
FreqDist({'France': 1, 'Tours.': 1, '(2,656': 1, 'stages': 1, 'years,': 1, 'tea
ms.': 1, 'placed': 1, 'third,': 1, 'behind': 1})
```

In [26]:

```
['My', 'fellow', 'citizens', ':', 'I', 'stand', 'here', ...]

---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
<ipython-input-26-cf4f9f33f6c4> in <module>
----> 1 print(inaugural.words(fileids = '2009-Obama.txt'))[0]

TypeError: 'NoneType' object is not subscriptable
```

In [ ]: