

```
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) ([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 [4]: from nltk.corpus import brown
```

```
In [5]: brown.categories()
```

```
Out[5]: ['adventure',
         'belles_lettres',
         'editorial',
         'fiction',
         'government',
         'hobbies',
         'humor',
         'learned',
         'lore',
         'mystery',
         'news',
         'religion',
         'reviews',
         'romance',
         'science_fiction']
```

```
In [6]: brown.words(categories = 'adventure')[:50]
```

```
Out[6]: ['Dan',
'Morgan',
'told',
'himself',
'he',
'would',
'forget',
'Ann',
'Turner',
'.',
'He',
'was',
'well',
'rid',
'of',
'her',
'.',
'He',
'certainly',
"didn't",
'want',
'a',
'wife',
'who',
'was',
'fickle',
'as',
'Ann',
'.',
'If',
'he',
'had',
'married',
'her',
',',
"he'd",
'have',
'been',
'asking',
'for',
'trouble',
'.',
'But',
'all',
'of',
'this',
'was',
'rationalization',
'.',
'Sometimes']
```

```
In [7]: from nltk.corpus import inaugural
```

```
In [8]: inaugural.fileids()
```

```
Out[8]: ['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 [9]: inaugural.words(fileids='1861-Lincoln.txt')[:20]
```

```
Out[9]: ['Fellow',  
        '-',  
        'Citizens',  
        'of',  
        'the',  
        'United',  
        'States',  
        ':',  
        'In',  
        'compliance',  
        'with',  
        'a',  
        'custom',  
        'as',  
        'old',  
        'as',  
        'the',  
        'Government',  
        'itself',  
        ',']
```

```
In [1]: from nltk.corpus import stopwords
```

```
In [4]: print(stopwords.words('english'))
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you'r  
e", "you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves',  
'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'i  
t', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves',  
'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those',  
'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'ha  
d', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but',  
'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'wit  
h', 'about', 'against', 'between', 'into', 'through', 'during', 'before', 'af  
ter', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off',  
'over', 'under', 'again', 'further', 'then', 'once', 'here', 'there', 'when',  
'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more', 'most',  
'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'th  
an', 'too', 'very', 's', 't', 'can', 'will', 'just', 'don', "don't", 'shoul  
d', "should've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren',  
"aren't", 'couldn', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn',  
"hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'might  
n', "mightn't", 'mustn', "mustn't", 'needn', "needn't", 'shan', "shan't", 'sh  
ouldn', "shouldn't", 'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'w  
ouldn', "wouldn't"]
```

```
In [5]: from nltk.corpus import stopwords

from nltk.tokenize import word_tokenize

example_sent = """This is a sample sentence,

                showing off the stop words filtration."""

stop_words = set(stopwords.words('english'))

word_tokens = word_tokenize(example_sent)
# converts the words in word_tokens to lower case and then checks whether
#they are present in stop_words or not

filtered_sentence = [w for w in word_tokens if not w.lower() in stop_words]
#with no lower case conversion

filtered_sentence = []

for w in word_tokens:

    if w not in stop_words:

        filtered_sentence.append(w)

print(word_tokens)

print(filtered_sentence)

['This', 'is', 'a', 'sample', 'sentence', ',', 'showing', 'off', 'the', 'sto',
p', 'words', 'filtration', '.']
['This', 'sample', 'sentence', ',', 'showing', 'stop', 'words', 'filtration',
'.']
```

```
In [6]: from nltk.corpus import webtext
```

```
In [7]: webtext.raw('grail.txt')
```

```
Out[7]: "SCENE 1: [wind] [clap clap clap] \nKING ARTHUR: Whoa there! [clap clap c
lop] \nSOLDIER #1: Halt! Who goes there?\nARTHUR: It is I, Arthur, son of
Uther Pendragon, from the castle of Camelot. King of the Britons, defeato
r of the Saxons, sovereign of all England!\nSOLDIER #1: Pull the other on
e!\nARTHUR: I am, ... and this is my trusty servant Patsy. We have ridde
n the length and breadth of the land in search of knights who will join me
in my court at Camelot. I must speak with your lord and master.\nSOLDIER
#1: What? Ridden on a horse?\nARTHUR: Yes!\nSOLDIER #1: You're using coco
nuts!\nARTHUR: What?\nSOLDIER #1: You've got two empty halves of coconut a
nd you're bangin' 'em together.\nARTHUR: So? We have ridden since the sno
ws of winter covered this land, through the kingdom of Mercea, through--\n
SOLDIER #1: Where'd you get the coconuts?\nARTHUR: We found them.\nSOLDIER
#1: Found them? In Mercea? The coconut's tropical!\nARTHUR: What do you
mean?\nSOLDIER #1: Well, this is a temperate zone.\nARTHUR: The swallow ma
y fly south with the sun or the house martin or the plover may seek warmer
climes in winter, yet these are not strangers to our land?\nSOLDIER #1: Ar
e you suggesting coconuts migrate?\nARTHUR: Not at all. They could be car
ried.\nSOLDIER #1: What? A swallow carrying a coconut?\nARTHUR: It could
grip it by the husk!\nSOLDIER #1: It's not a question of where he grips i
```

```
In [ ]:
```