

21MIC0087 Hands-on-session 1

```
In [ ]: import nltk
        #nltk.download()

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

```
In [1]: 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 [2]: from nltk.corpus import brown
```

```
In [3]: brown.categories()
```

```
Out[3]: ['adventure',
         'belles_lettres',
         'editorial',
         'fiction',
         'government',
         'hobbies',
         'humor',
         'learned',
         'lore',
         'mystery',
         'news',
         'religion',
         'reviews',
         'romance',
         'science_fiction']
```

```
In [4]: brown.words(categories='adventure')[:50]
```

```
Out[4]: ['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 [5]: from nltk.corpus import inaugural
```

```
In [6]: inaugural.fileids()
```

```
Out[6]: ['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 [8]: inaugural.words(fileids='1861-Lincoln.txt')[:20]
```

```
Out[8]: ['Fellow',  
        '-',  
        'Citizens',  
        'of',  
        'the',  
        'United',  
        'States',  
        ':',  
        'In',  
        'compliance',  
        'with',  
        'a',  
        'custom',  
        'as',  
        'old',  
        'as',  
        'the',  
        'Government',  
        'itself',  
        ',']
```