

WINTER SEMESTER 2022-23

CSE4022-NATURAL LANGUAGE PROCESSING

NAME: Kancharla Bhavana

REG NO.: 20BCE2732

SLOT: E1+TE1

VENUE: SJT626

TASK-1

1) Installing NLTK

```
In [1]: pip install nltk
```

```
Requirement already satisfied: nltk in e:\anaconda\lib\site-packages (3.7)
Requirement already satisfied: joblib in e:\anaconda\lib\site-packages (from nltk) (1.1.0)
Requirement already satisfied: regex>=2021.8.3 in e:\anaconda\lib\site-packages (from nltk) (2022.7.9)
Requirement already satisfied: tqdm in e:\anaconda\lib\site-packages (from nltk) (4.64.1)
Requirement already satisfied: click in e:\anaconda\lib\site-packages (from nltk) (8.0.4)
Requirement already satisfied: colorama in e:\anaconda\lib\site-packages (from click->nltk) (0.4.5)
Note: you may need to restart the kernel to use updated packages.
```

```
In [ ]:
```

2) Downloading NLTK

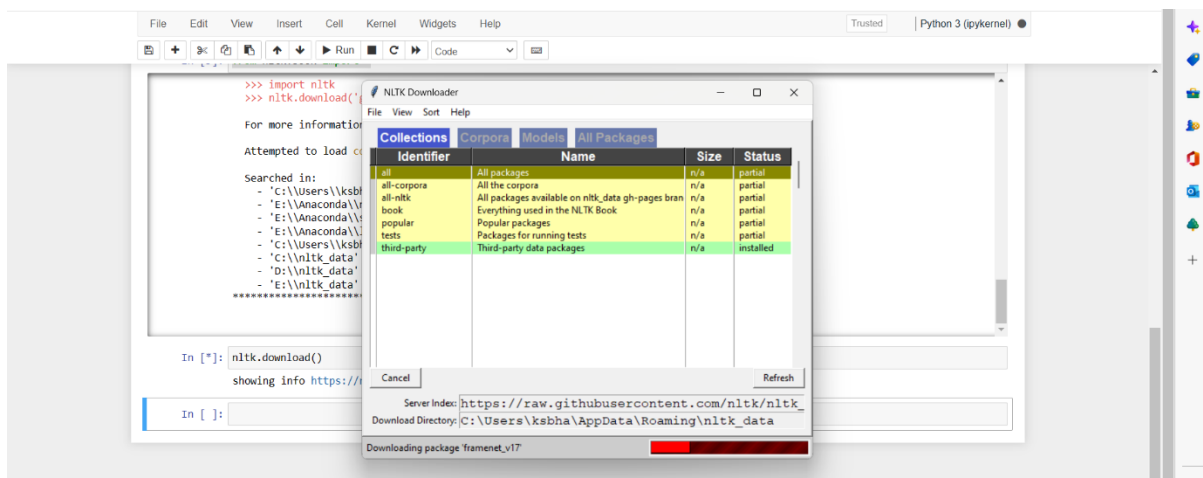
```
In [2]: import nltk
```

```
In [3]: nltk.download()
```

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

```
Out[3]: True
```

```
In [ ]:
```



NLTK Downloader

File
View
Sort
Help

Collections
Corpora
Models
All Packages

Identifier	Name	Size	Status
all	All packages	n/a	partial
all-corpora	All the corpora	n/a	partial
all-nltk	All packages available on nltk_data gh-pages bran	n/a	partial
book	Everything used in the NLTK Book	n/a	partial
popular	Popular packages	n/a	partial
tests	Packages for running tests	n/a	installed
third-party	Third-party data packages	n/a	installed

Cancel
Refresh

Server Index:

Download Directory:

Downloading package 'wordnet31'

NLTK Downloader

File
View
Sort
Help

Collections
Corpora
Models
All Packages

Identifier	Name	Size	Status
all	All packages	n/a	installed
all-corpora	All the corpora	n/a	installed
all-nltk	All packages available on nltk_data gh-pages bran	n/a	installed
book	Everything used in the NLTK Book	n/a	installed
popular	Popular packages	n/a	installed
tests	Packages for running tests	n/a	installed
third-party	Third-party data packages	n/a	installed

Download
Refresh

Server Index:

Download Directory:

Finished downloading collection 'all'.

```
In [10]: import nltk
```

```
In [11]: nltk.download('nps_chat')
nltk.download('webtext')
nltk.download('popular')
nltk.download('brown')
from nltk.book import *
```

```
[nltk_data] Downloading package nps_chat to
[nltk_data] C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] Package nps_chat is already up-to-date!
[nltk_data] Downloading package webtext to
[nltk_data] C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] Package webtext is already up-to-date!
[nltk_data] Downloading collection 'popular'
[nltk_data] |
[nltk_data] | Downloading package cmudict to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package cmudict is already up-to-date!
[nltk_data] | Downloading package gazetteers to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package gazetteers is already up-to-date!
[nltk_data] | Downloading package genesis to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package genesis is already up-to-date!
[nltk_data] | Downloading package gutenberg to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package gutenberg is already up-to-date!
[nltk_data] | Downloading package inaugural to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package inaugural is already up-to-date!
[nltk_data] | Downloading package movie_reviews to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
```

```
[nltk_data] | Package movie_reviews is already up-to-date!
[nltk_data] | Downloading package names to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package names is already up-to-date!
[nltk_data] | Downloading package shakespeare to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package shakespeare is already up-to-date!
[nltk_data] | Downloading package stopwords to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package stopwords is already up-to-date!
[nltk_data] | Downloading package treebank to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package treebank is already up-to-date!
[nltk_data] | Downloading package twitter_samples to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package twitter_samples is already up-to-date!
[nltk_data] | Downloading package omw to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package omw is already up-to-date!
[nltk_data] | Downloading package omw-1.4 to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package omw-1.4 is already up-to-date!
[nltk_data] | Downloading package wordnet to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet is already up-to-date!
[nltk_data] | Downloading package wordnet2021 to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet2021 is already up-to-date!
[nltk_data] | Downloading package wordnet31 to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet31 is already up-to-date!
[nltk_data] | Downloading package wordnet_ic to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet_ic is already up-to-date!
[nltk_data] | Downloading package words to
```

```
[nltk_data] | Downloading package wordnet_ic to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet_ic is already up-to-date!
[nltk_data] | Downloading package words to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package words is already up-to-date!
[nltk_data] | Downloading package maxent_ne_chunker to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package maxent_ne_chunker is already up-to-date!
[nltk_data] | Downloading package punkt to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package punkt is already up-to-date!
[nltk_data] | Downloading package snowball_data to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package snowball_data is already up-to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger to
[nltk_data] | C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] | Package averaged_perceptron_tagger is already up-
[nltk_data] | to-date!
[nltk_data] |
[nltk_data] Done downloading collection popular
[nltk_data] Downloading package brown to
[nltk_data] C:\Users\ksbha\AppData\Roaming\nltk_data...
[nltk_data] Package brown is already up-to-date!
```

```
In [ ]:
```

3) Importing NLTK books

```
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 [ ]:
```

4) Brown Corpus

```
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 [ ]:
```

```
In [4]: brown.words(categories='news')[:10]
```

```
Out[4]: ['The',
         'Fulton',
         'County',
         'Grand',
         'Jury',
         'said',
         'Friday',
         'an',
         'investigation',
         'of']
```

```
In [ ]:
```

5) Inaugural Corpus

```
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',
```

```
        '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 [ ]:
```

```
In [7]: inaugural.words(fileids='1861-Lincoln.txt')[:10]
```

```
Out[7]: ['Fellow',  
        '-',  
        'citizens',  
        'of',  
        'the',  
        'United',  
        'States',  
        ':',  
        'In',  
        'compliance']
```

```
In [ ]:
```

```
In [8]: inaugural.words(fileids='2013-Obama.txt')[:10]
```

```
Out[8]: ['Thank', 'you', '.', 'Thank', 'you', 'so', 'much', '.', 'Vice', 'President']
```

```
In [ ]:
```