

Winter Semester 2022 – 2023
CSE4022 – Natural Language Processing
Task – 1

Name: T Harshitha Deepthi

Date: 16 – 12 – 22

Reg no: 20BCE2019

Slot: E1 + TE1

Venue: SJT626

1. Intsalling nltk

```
In [3]: pip install nltk
```

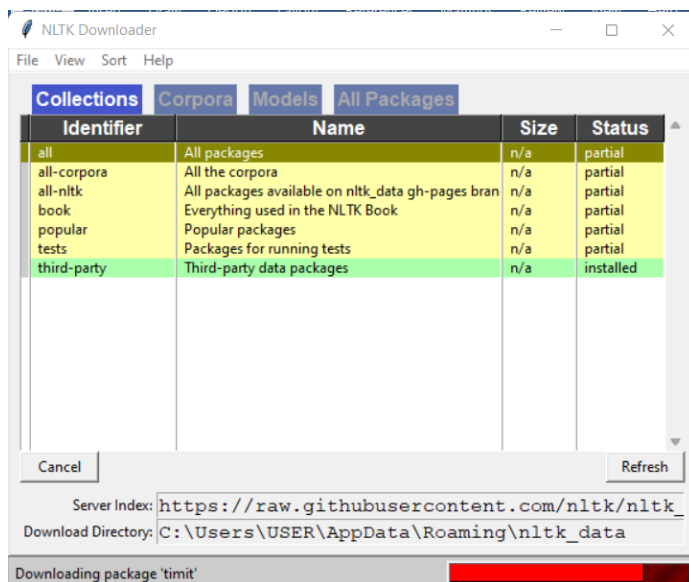
```
Requirement already satisfied: nltk in c:\users\user\anaconda3\lib\site-packages (3.5)Note: you may need to restart the kernel  
to use updated packages.  
Requirement already satisfied: regex in c:\users\user\anaconda3\lib\site-packages (from nltk) (2020.6.8)  
Requirement already satisfied: joblib in c:\users\user\anaconda3\lib\site-packages (from nltk) (0.16.0)  
Requirement already satisfied: click in c:\users\user\anaconda3\lib\site-packages (from nltk) (7.1.2)  
Requirement already satisfied: tqdm in c:\users\user\anaconda3\lib\site-packages (from nltk) (4.47.0)
```

2. Downloading nltk

```
In [4]: import nltk
```

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

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



```
In [4]: import nltk
```

```
In [6]: nltk.download()
```

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

```
Out[6]: True
```

3. Importing nltk.books

```
In [7]: import nltk
```

```
In [8]: 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: Personalads Corpus
text9: The Man Who Was Thursday by G . K . Chesterton 1908
```

4. Brown corpus

```
In [9]: from nltk.corpus import brown
```

```
In [10]: brown.categories()
```

```
Out[10]: ['adventure',
          'belles_lettres',
          'editorial',
          'fiction',
          'government',
          'hobbies',
          'humor',
          'learned',
          'lore',
          'mystery',
          'news',
          'religion',
          'reviews',
          'romance',
          'science_fiction']
```

```
In [11]: brown.words(categories='news')[:10]
```

```
Out[11]: ['The',
          'Fulton',
          'County',
          'Grand',
          'Jury',
          'said',
          'Friday',
          'an',
          'investigation',
          'of']
```

5. Inaugural corpus

```
In [13]: from nltk.corpus import inaugural
```

```
In [14]: inaugural.fileids()
```

```
Out[14]: ['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 [15]: inaugural.words(fileids='1861-Lincoln.txt')[:10]
```

```
Out[15]: ['Fellow',  
         '...',  
         'Citizens',  
         'of',  
         'the',  
         'United',  
         'States',  
         ':',  
         'In',  
         'compliance']
```

```
In [16]: inaugural.words(fileids='2013-Obama.txt')[:10]
```

```
Out[16]: ['Thank', 'you', '.', 'Thank', 'you', 'so', 'much', '.', 'Vice', 'President']
```

Downloading some packages:

```
In [12]: 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\USER\AppData\Roaming\nltk_data...  
[nltk_data] Package nps_chat is already up-to-date!  
[nltk_data] Downloading package webtext to  
[nltk_data] C:\Users\USER\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\USER\AppData\Roaming\nltk_data...  
[nltk_data] | Package cmudict is already up-to-date!  
[nltk_data] | Downloading package gazetteers to  
[nltk_data] | C:\Users\USER\AppData\Roaming\nltk_data...  
[nltk_data] | Package gazetteers is already up-to-date!  
[nltk_data] | Downloading package genesis to  
[nltk_data] | C:\Users\USER\AppData\Roaming\nltk_data...  
[nltk_data] | Package genesis is already up-to-date!  
[nltk_data] | Downloading package gutenberg to  
[nltk_data] | C:\Users\USER\AppData\Roaming\nltk_data...  
[nltk_data] | Package gutenberg is already up-to-date!  
[nltk_data] | Downloading package inaugural to  
[nltk_data] | C:\Users\USER\AppData\Roaming\nltk_data...  
[nltk_data] | Package inaugural is already up-to-date!  
[nltk_data] | Downloading package movie_reviews to  
[nltk_data] | C:\Users\USER\AppData\Roaming\nltk_data...  
[nltk_data] | Package movie_reviews is already up-to-date!
```

```
[nltk_data] | Downloading package maxent_ne_chunker to  
[nltk_data] | C:\Users\USER\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\USER\AppData\Roaming\nltk_data...  
[nltk_data] | Package punkt is already up-to-date!  
[nltk_data] | Downloading package snowball_data to  
[nltk_data] | C:\Users\USER\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\USER\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\USER\AppData\Roaming\nltk_data...  
[nltk_data] Package brown is already up-to-date!
```