Natural Language Processing

Practicals

Practical 1

Aim:a.Install NLTK

```
# 01_A Install nltk
#https://www.tutorialspoint.com/natural_language_toolkit/natural_language_toolkit_getting_
# go to terminal window and type :
#conda install -c anaconda nltk
import nltk
nltk.download()
```

```
       [ ] sentiwordnet........ SentiWordNet
       [ ] shakespeare......... Shakespeare XML Corpus Sample
       [ ] sinica_treebank..... Sinica Treebank Corpus Sample
       [ ] smultron............ SMULTRON Corpus Sample
       [ ] snowball_data....... Snowball Data
       [ ] spanish_grammars.... Grammars for Spanish
       [ ] state_union......... C-Span State of the Union Address Corpus
       [ ] stopwords........... Stopwords Corpus
       [ ] subjectivity........ Subjectivity Dataset v1.0
       [ ] swadesh............. Swadesh Wordlists
       [ ] switchboard......... Switchboard Corpus Sample
       [ ] tagsets............. Help on Tagsets
       [ ] timit............... TIMIT Corpus Sample
       [ ] toolbox............. Toolbox Sample Files
       [ ] treebank............ Penn Treebank Sample
       [ ] twitter_samples..... Twitter Samples
     Hit Enter to continue:
       [ ] udhr2............... Universal Declaration of Human Rights Corpus
                                (Unicode Version)
       [ ] udhr................ Universal Declaration of Human Rights Corpus
       [ ] unicode_samples..... Unicode Samples
       [ ] universal_tagset.... Mappings to the Universal Part-of-Speech Tagset
       [ ] universal_treebanks_v20 Universal Treebanks Version 2.0
       [ ] vader_lexicon....... VADER Sentiment Lexicon
       [ ] verbnet3............ VerbNet Lexicon, Version 3.3
       [ ] verbnet............. VerbNet Lexicon, Version 2.1
       [ ] webtext............. Web Text Corpus
       [ ] wmt15_eval.......... Evaluation data from WMT15
       [ ] word2vec_sample..... Word2Vec Sample
       [ ] wordnet2021......... Open English Wordnet 2021
       [ ] wordnet31........... Wordnet 3.1
       [ ] wordnet............. WordNet
       [ ] wordnet_ic.......... WordNet-InfoContent
       [ ] words............... Word Lists
       [ ] ycoe................ York-Toronto-Helsinki Parsed Corpus of Old
                                English Prose
```

```
Collections:
  [ ] all-corpora......... All the corpora
Hit Enter to continue:
  [ ] all-nltk............ All packages available on nltk_data gh-pages
                           branch
  [ ] all................. All packages
  [ ] book................ Everything used in the NLTK Book
  [ ] popular............. Popular packages
  [ ] tests............... Packages for running tests
  [ ] third-party......... Third-party data packages

([*] marks installed packages)

Download which package (l=list; x=cancel)?
  Identifier> x

---------------------------------------------------------------------------
    d) Download   l) List    u) Update   c) Config   h) Help   q) Quit
---------------------------------------------------------------------------
Downloader> q
True
```

.

```
import nltk
print('nltk imported')
```

```
nltk imported
```

.

```
from nltk.stem import PorterStemmer
word_stemmer = PorterStemmer()
print(word_stemmer.stem('coding'))
print(word_stemmer.stem('eating'))
```

```
code
eat
```

.

.

b.Convert the given text to speech

.

```
pip install gTTS
```

```
Requirement already satisfied: gTTS in /usr/local/lib/python3.7/dist-packages (2.2.3
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from g
Requirement already satisfied: click in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-pa
```

.

.

```
"""
01_B Convert the given text to speech
"""

from gtts import gTTS

# This module is imported so that we can
# play the converted audio
import os

# The text that you want to convert to audio
mytext = 'Hi Dr. Mahendra K here'

# Language in which you want to convert
language = 'en'

# Passing the text and language to the engine,
# here we have marked slow=False. Which tells
# the module that the converted audio should
# have a high speed
myobj = gTTS(text=mytext, lang=language, slow=False)

# Saving the converted audio in a mp3 file named
# welcome
myobj.save("welcome.mp3")

# Playing the converted file
os.system("welcome.wav")
```

```
32512
```

.

c.Convert audio file Speech to Text

```
pip install SpeechRecognition
```

```
Requirement already satisfied: SpeechRecognition in /usr/local/lib/python3.7/dist-pa
```

.

.

```
import speech_recognition as sr

AUDIO_FILE = ("harvard.wav")

# use the audio file as the audio source
r = sr.Recognizer()
with sr.AudioFile(AUDIO_FILE) as source:
  #reads the audio file. Here we use record instead of
  #listen
  audio = r.record(source)

try:
  print("The audio file contains: \n" + r.recognize_google(audio))

except sr.UnknownValueError:
  print("Google Speech Recognition could not understand audio")

except sr.RequestError as e:
  print("Could not request results from Google Speech Recognition service; {0}".format(e))
```

```
The audio file contains:
the stale smell of old beer lingers it takes heat to bring out the odor a cold dip r
```