Tanay Jhawar

BE4 Roll No:22

EXPERIMENT 5

AIM: Implement Morphological Analysis in Python

SOURCE CODE:

```
import nltk
from nltk.stem.wordnet import WordNetLemmatizer
from nltk.corpus import wordnet
from nltk.tokenize import sent_tokenize, word_tokenize
from nltk.corpus import stopwords
from textblob import TextBlob

#Morphological Analysis
def get_wordnet_pos(word):
   tag = nltk.pos_tag([word])[0][1][0].upper()
   tag_dict = {"J":wordnet.ADJ,
               "N":wordnet.NOUN,
               "V":wordnet.VERB,
               "R":wordnet.ADV}
   return tag_dict.get(tag,wordnet.NOUN)

f=open("Tanay_Exp5.txt")
stop_words = set(stopwords.words("English"))
bad_chars = (';',':','!', "*",'<','>','#','?','@','p',',','.','(',')','eos')

for line in f:
   words = word_tokenize(line)
   print("Word Tokenization\n")
   print(words)
```

```
    print("\n")
    #stop word and filterization
    without_stop_words = [word for word in words if not word in stop_words]
    ws = list(filter(lambda i: i not in bad_chars,without_stop_words))
    print("Filtered Words \n")
    print(ws)
    print("\n")
    ws_tag = nltk.pos_tag(ws)

wordtaglength = len(ws_tag)

#Lemmatizer
l = WordNetLemmatizer()

print("WORD","\t\t\t","ROOT WORD(MORPHENE)","\t\t","TAG")
for i in range(wordtaglength):
        print(ws[i].ljust(9),"\t\t",l.lemmatize(ws[i],get_wordnet_pos(ws[i])).ljust(9),"\t\t",ws_tag[i])
```

OUTPUT:


Python 3.7.9 Shell
File Edit Shell Debug Options Window Help
Python 3.7.9 (tags/v3.7.9:13c94747c7, Aug 17 2020, 16:30:00) [MSC v.1900 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
========== RESTART: C:\Users\Dell\Desktop\SEM8\NLP\Exp5.py ==========
Word Tokenization
['Football', 'or', 'soccer', ',', 'is', 'a', 'team', 'sport', '.', 'It', 'is', 'played', 'by', 'approximately', '250', 'million', 'players', 'in', 'over', '200', 'countries', 'and', 'dependencies', '.', 'Tanay', 'plays', 'football', '.']
Filtered Words
['Football', 'soccer', 'team', 'sport', 'It', 'played', 'approximately', '250', 'million', 'players', '200', 'countries', 'dependencies', 'Tanay', 'plays', 'football']
WORD ROOT WORD(MORPHENE) TAG
Football Football ('Football', 'NN')
soccer soccer ('soccer', 'NN')
team team ('team', 'NN')
sport sport ('sport', 'NN')
It It ('It', 'PRP')
played played ('played', 'VBD')
approximately approximately ('approximately', 'RB')
250 250 ('250', 'CD')
million million ('million', 'CD')
players player ('players', 'NNS')
200 200 ('200', 'CD')
countries country ('countries', 'NNS')
dependencies dependency ('dependencies', 'NNS')
Tanay Tanay ('Tanay', 'NNP')
plays play ('plays', 'VBZ')
football football ('football', 'NN')
>>>