

▼ Install PyTorch & INLTK

```
!pip install torch==1.3.1+cpu -f https://download.pytorch.org/whl/torch\_stable.html  
!pip install inltk
```



```
Looking in links: https://download.pytorch.org/whl/torch\_stable.html
Requirement already satisfied: torch==1.3.1+cpu in /usr/local/lib/python3.6/dist-packages
Requirement already satisfied: numpy in /usr/local/lib/python3.6/dist-packages (from
Requirement already satisfied: inltk in /usr/local/lib/python3.6/dist-packages (0.8.1
```

```
import warnings
```

```
warnings.filterwarnings('ignore')
```

```
Requirement already satisfied: torch==1.3.1+cpu in /usr/local/lib/python3.6/dist-packages
Requirement already satisfied: numpy in /usr/local/lib/python3.6/dist-packages (from
```

▼ Import Library

```
Requirement already satisfied: fastai==1.0.57 in /usr/local/lib/python3.6/dist-packages
```

```
from inltk.inltk import setup
```

```
setup('gu')
```

```
☞ Done!
```

```
Requirement already satisfied: aiohttp>=3.5.4 in /usr/local/lib/python3.6/dist-packages
```

▼ Task-1. - Tokenization

```
Requirement already satisfied: numpy>=1.15 in /usr/local/lib/python3.6/dist-packages
```

```
from inltk.inltk import tokenize
```

```
text = '''ગીર રાષ્ટ્રીય ઉદ્યાન અને ગીર અભયારણ્ય (જે "ગીરનું જંગલ" કે "સાસણ-ગીર" તરીકે પણ ઓળખાય છે)
ગુજરાતમાં આવેલું જંગલ અને વન્યજીવન અભયારણ્ય છે. તેની સ્થાપના ૧૯૬૫માં કરવામાં આવી હતી,
તે કુલ ૧,૪૧૨ ચો.કી.મી. ના વિસ્તારમાં ફેલાયેલ છે.'''
```

```
tokenize(text, 'gu')
```

```
☞
```

```

'રાષ્ટ્રીય',
'ઉદ્યાન',
'અને',
'ગીર',
'અભયારણ્ય',
',',
'(',
'જે',
',',
',',
'ગીર',
'નું',
'જંગલ',
',',
',',
'કે',
',',
',',
'સ',
',',
'ાસણ',
',',
',',
'ગીર',
',',
',',
'તરીકે',
',',
'પણ',
',',
'ઓળખાય',
',',
'છે',
',',
',',
'ગુજરાત',
',',
'માં',
',',
'આવેલું',
',',
'જંગલ',
',',
'અને',
',',
'વન્યજીવન',

```

▼ Task-2. - Embedding Vectors

```

_સ્યાપના',

from inltk.inltk import get_embedding_vectors

text = "ભારત મારો દેશ છે"
vectors = get_embedding_vectors(text, 'gu')
print("Length of Vector : ", len(vectors))
print("Shape of Vector : ", vectors[0].shape)
print(vectors)

↳ Length of Vector : 4
Shape of Vector : (400,)
[array([-0.118378, -0.206994, -0.375069,  0.188438, ...,  0.056597,  0.574319,  0.064

',',

from inltk.inltk import get_sentence_encoding
get_sentence_encoding("ભારત મારો દેશ છે.", 'gu')

↳ array([-0.018772, -0.018323,  0.021537, -0.046115, ...,  0.211418, -0.159044, -0.0256

_ૉ',

```

▼ Task-3. - Predict Next 'n' words

```
from nltk.inltk import predict_next_words
text = '''ગીર રાષ્ટ્રીય ઉદ્યાન અને ગીર અભયારણ્ય (જે "ગીરનું જંગલ" કે "સાસણ-ગીર" તરીકે પણ ઓળખાય છે) ગુજરાતમાં આવેલું જંગલ અને વન્યજીવન અભયારણ્ય છે. તેની સ્થાપના ૧૯૬૫માં કરવામાં આવી હતી, તે કુલ ૧,૪૧૨ ચો.કી.મી. ના વિસ્તારમાં ફેલાયેલ છે '''
predict_next_words(text , 10, 'gu')
```

☞ 'ગીર રાષ્ટ્રીય ઉદ્યાન અને ગીર અભયારણ્ય (જે "ગીરનું જંગલ" કે "સાસણ-ગીર" તરીકે પણ ઓળખાય છે) \nગુજરાતમાં આવેલું જંગલ અને વન્યજીવન અભયારણ્ય છે. તેની સ્થાપના ૧૯૬૫માં કરવામાં આવી હતી, \nતે કુલ ૧.૪૧૨ ચો કી મી ના વિસ્તારમાં ફેલાયેલ છે અને લગભગ ૮.૭ .૩૩ ચોરસ કિલોમીટર જેટલો વિ

▼ Task-4. - Remove Other languages

```
from nltk.inltk import remove_foreign_languages
```

```
text = '''Trainees were awarded a certificate for successful completion of the program.
કાર્યક્રમની સફળ કામગીરી માટે તાલીમાર્થીઓને પ્રમાણપત્ર એનાયત કરાયા હતા.'''
```

```
remove_foreign_languages(text, 'gu')
```

```
#If we take Gujarati as a host Language then it should remove words
#which are in Language other than gujarati(English in this case)
#But here it is not working.
```

☞

```

['_T',
 'ra',
 'ine',
 'es',
 '_were',
 '_award',
 '_']

setup('en')

from nltk.nltk import remove_foreign_languages

text = '''Trainees were awarded a certificate for successful completion of the program.
કાર્યક્રમની સફળ કામગીરી માટે તાલીમાર્થીઓને પ્રમાણપત્ર એનાયત કરાયા હતા.'''

remove_foreign_languages(text, 'en')

#If we take English as a host Language then it should remove words which are in
#Language other than English(Gujarati in this case) and it mark as 'xxunk' in the result.
#For English it remove other than english words and execute properly.

☞ Done!
['trainees',
 'were',
 'awarded',
 'a',
 'certificate',
 'for',
 'successful',
 'completion',
 'of',
 'the',
 'program',
 '.',
 '\n ',
 'xxunk',
 'xxunk',
 'xxunk',
 'xxunk',
 'xxunk',
 'xxunk',
 'xxunk',
 'xxunk',
 'xxunk',
 'xxunk']

```

▼ Task-5. - Sentence Similarity

```

from nltk.nltk import get_sentence_similarity
sent_1 = "તમે ફરીથી તે બોલી શકો છો?"
sent_2 = "શું તમે ગુજરાતી બોલી શકો છો"

print("\nSimilarity Between Given Sentences\n")
get_sentence_similarity(sent_1, sent_2, 'gu')

```

☞

Similarity Between Given Sentences

▼ Task-6. - Get Similar Sentences

```
from nltk.nltk import get_similar_sentences  
sentence = "હું આજે ખુબ ખુશ છું"  
get_similar_sentences(sentence, 5, 'gu', degree_of_aug = 0.2)
```

```
↳ ['હું અત્યારે ખુબ ખુશ છું',  
    'હું આજે ખુબજ ખુશ છું',  
    'હું આજે ખુબજ ખુશ છું',  
    'હું આજકાલ ખુબ ખુશ છું',  
    'તમે આજે ખુબ ખુશ છું']
```

Could not connect to the reCAPTCHA service. Please check your internet connection and reload to get a reCAPTCHA challenge.