

CSE6060

Statistical Natural Language Processing

NLP - தமிழ்

Name : Kavianand G

Reg. No. : 19MAI0050

Date : 27 – June – 2020

Indic NLP Library

The library provides the following functionalities:

- Text Normalization
- Script Information
- Tokenization
- Word Segmentation
- Script Conversion
- Romanization
- Indicization
- Transliteration
- Translation

In [1]:

```
1 # The path to the local git repo for Indic NLP Library
2 INDIC_NLP_LIB_HOME=r"C:\Users\kavianand\Documents\src\indic_nlp_library"
3
4 # The path to the local git repo for Indic NLP Resources
5 INDIC_NLP_RESOURCES=r"C:\Users\kavianand\Documents\src\indic_nlp_resources"
```

In [2]:

```
1 import sys
2 sys.path.append(r'{}\src'.format(INDIC_NLP_LIB_HOME))
```

In [3]:

```
1 from indicnlp import common
2 common.set_resources_path(INDIC_NLP_RESOURCES)
```

In [4]:

```
1 from indicnlp import loader
2 loader.load()
```

In [5]:

```
1 data="""தமிழ் இலக்கியம் இரண்டாயிரம் ஆண்டுகளுக்கு மேலான தொடர்ச்சி கொண்
2
3 print(data)
```

"தமிழ் இலக்கியம் இரண்டாயிரம் ஆண்டுகளுக்கு மேலான தொடர்ச்சி கொண்ட உலகின் சிறந்த இலக்கியங்களில் ஒன்று. வாழ்வின் பல்வேறு கூறுகளை தமிழ் இலக்கியங்கள் இயம்புகின்றன. தமிழ் இலக்கியத்தில் வெண்பா, குறள், புதுக்கவிதை, கட்டுரை, பழமொழி, தொண்ணூற்றாறு வகை சிற்றிலக்கியங்கள் என பல வடிவங்கள் உள்ளன. தமிழில் வாய்மொழி இலக்கியங்களும் முக்கிய இடம் வகிக்கின்றன.

Text Normalization

In [6]:

```
1 from indicnlp.normalize.indic_normalize import IndicNormalizerFactory
2
3 input_text= data
4
5 factory=IndicNormalizerFactory()
6 normalizer=factory.get_normalizer("ta")
7 output_text=normalizer.normalize(input_text)
8
9 print(input_text)
10 print()
11
12 print('Before normalization')
13 print(' '.join([ c for c in input_text ] ))
14 print('Length: {}'.format(len(input_text)))
15 print()
16 print('After normalization')
17 print(' '.join([ c for c in output_text ] ))
18 print('Length: {}'.format(len(output_text)))
19
```

"தமிழ் இலக்கியம் இரண்டாயிரம் ஆண்டுகளுக்கு மேலான தொடர்ச்சி கொண்ட உலகின் சிறந்த இலக்கியங்களில் ஒன்று. வாழ்வின் பல்வேறு கூறுகளை தமிழ் இலக்கியங்கள் இயம்புகின்றன. தமிழ் இலக்கியத்தில் வெண்பா, குறள், புதுக்கவிதை, கட்டுரை, பழமொழி, தொண்ணூற்றாறு வகை சிற்றிலக்கியங்கள் என பல வடிவங்கள் உள்ளன. தமிழில் வாய்மொழி இலக்கியங்களும் முக்கிய இடம் வகிக்கின்றன.

Before normalization

" த ம ி ழ ி இ ல க ி க ி ய ம ி இ ர ண ஂ ட ா ய ி ர ம ி ஆ ண ஂ ட ஁ க ள ஁ க ி க ி ம ே ல ா ன த ி ட ர ி ச ி க ி ண ஂ ட உ ல க ி ன ி ச ி ற ன ஂ த இ ல க ி க ி ய ன ி க ள ி ல ி ஒ ன ஂ ற ஁ . வ ா ழ ி வ ி ன ி ப ல ி வ ே ற ஁ க ி ற ஁ க ள ன ை த ம ி ழ ி இ ல க ி க ி ய ன ி க ள ி இ ய ம ி ப ஁ க ி ன ஂ ற ன . த ம ி ழ ி இ ல க ி க ி ய த ி த ி ல ி வ ஂ ண ஂ ப ா , க ி ற ள ி , ப ஁ த ஁ க ி க வ ி த ன ை , க ட ஂ ட ஁ ர ன ை , ப ழ ம ி ழ ி , த ி ண ஂ ண ி ற ி ற ஁ வ க ன ை ச ி ற ி ற ி ல க ி க ி ய ன ி க ள ி என ப ல வ ட ி வ ன ி க ள ி உ ள ன ன . த ம ி ழ ி ல ி வ ா ய ி ம ி ழ ி இ ல க ி க ி ய ன ி க ள ி ம ி ம ஁ க ி க ி ய இ ட ம ி வ க ி க ி க ி ன ஂ ற ன .

Length: 338

After normalization

" த ம ி ழ ி இ ல க ி க ி ய ம ி இ ர ண ஂ ட ா ய ி ர ம ி ஆ ண ஂ ட ஁ க ள ி க ி க ி ம ே ல ா ன த ி ட ர ி ச ி க ி ண ஂ ட உ ல க ி ன ி ச ி ற ன ஂ த இ ல க ி க ி ய ன ி க ள ி ல ி ஒ ன ஂ ற ஁ . வ ா ழ ி வ ி ன ி ப ல ி வ ே ற ஁ க ி ற ஁ க ள ன ை த ம ி ழ ி இ ல க ி க ி ய ன ி க ள ி இ ய ம ி ப ஁ க ி ன ஂ ற ன . த ம ி ழ ி இ ல க ி க ி ய த ி த ி ல ி வ ஂ ண ஂ ப ா , க ி ற ள ி , ப ஁ த ஁ க ி க வ ி த ன ை , க ட ஂ ட ஁ ர ன ை , ப ழ ம ி ழ ி , த ி ண ஂ ண ி ற ி ற ஁ வ க ன ை ச ி ற ி ற ி ல க ி க ி ய ன ி க ள ி என ப ல வ ட ி வ ன ி க ள ி உ ள ன ன . த ம ி ழ ி ல ி வ ா ய ி ம ி ழ ி இ ல க ி க ி ய ன ி க ள ி ம ி ம ஁ க ி க ி ய இ ட ம ி வ க ி க ி க ி ன ஂ ற ன .

Length: 338

Sentence Tokenization / Sentence Splitter

In [7]:

```
1 from indicnlp.tokenize import sentence_tokenize
2
3 string = data
4 sentences=sentence_tokenize.sentence_split(string, lang='ta')
5 for t in sentences:
6     print("\n",t)
```

"தமிழ் இலக்கியம் இரண்டாயிரம் ஆண்டுகளுக்கு மேலான தொடர்ச்சி கொண்ட உலகின் சிறந்த இலக்கியங்களில் ஒன்று.

வாழ்வின் பல்வேறு கூறுகளை தமிழ் இலக்கியங்கள் இயம்புகின்றன.

தமிழ் இலக்கியத்தில் வெண்பா, குறள், புதுக்கவிதை, கட்டுரை, பழமொழி, தொண்ணூற்றாறு வகை சிற்றிலக்கியங்கள் என பல வடிவங்கள் உள்ளன.

தமிழில் வாய்மொழி இலக்கியங்களும் முக்கிய இடம் வகிக்கின்றன.

In [8]:

```
1 data = ""
2 தென்னிந்தியக் கட்டிடக்கலையில் ஒரு திருப்புமுனையாக அமைந்து அதன் வளர்ச்சியி
3
4 பல்லவ வம்சத்தினர் தமிழகத்தின் வடபகுதியில் கி.பி. 4 ஆம் நூற்றாண்டு தொடக்கம்
5
6 பிற்காலப் பல்லவர்கள் காலத்திலேயே தமிழ் நாட்டில் இந்து சமய மறுமலர்ச்சியும், சி
7
8 பல்லவர் கட்டிடக்கலையை இரண்டு கட்டங்களாகப் பிரிப்பதுண்டு. முதல் கட்டம் சும
9
10 இவற்றுள் முதல் கட்டத்தில் கட்டப்பட்டவை மகேந்திரன் கட்டிடங்கள், மாமல்லன் கட்டிட
11
12 இரண்டாம் கட்டமும், நரசிம்மன் கட்டிடங்கள், நந்திவர்மன் கட்டிடங்கள் என இரண்டு
13
14 பழங்காலத்தில் தமிழ் நாட்டில் கட்டிடங்கள் அழியக்கூடிய கட்டிடப்பொருட்களான மரப
15
16 தமிழ் நாட்டுக்கு வடக்கே ஏற்கனவே முகம் காட்டத் தொடங்கியிருந்த திராவிடக் கட்டிட
17
18
19 பல்லவர்கள் அமைத்த குடைவரைகள் தமிழ் நாட்டில், மண்டகப்பட்டு, பல்லாவரம், ம
20
21 பஞ்சபாண்டவர் ரதங்கள் - மாமல்லபுரம்
22
23
24 தம்புள்ளையில் அமைந்துள்ள நாலந்த சிலை மண்டபம் இலங்கை கட்டிடக்கலையில் ப
25
26
27
28 ""
```

In [9]:

```
1 from indicnlp.tokenize import sentence_tokenize
2
3 string = data
4 sentences=sentence_tokenize.sentence_split(string, lang='ta')
5 for t in sentences:
6     print("\n",t)
```

தென்னிந்தியக் கட்டிடக்கலையில் ஒரு திருப்புமுனையாக அமைந்து அதன் வளர்ச்சியில் ஆகக் கூடிய தாக்கத்தை ஏற்படுத்தியது பல்லவர் காலமே எனில் அது மிகையல்ல.

இப்பகுதியில் கி. பி. 6 ஆம் நூற்றாண்டளவில் ஆரம்பித்து வளர்ச்சியடைந்த திராவிடக் கட்டிடக்கலையின் ஆரம்பம் இந்தக் காலமே.

பல்லவ வம்சத்தினர் தமிழகத்தின் வடபகுதியில் கி. பி. 4 ஆம் நூற்றாண்டு தொடக்கம் கி.

பி. 10 ஆம் நூற்றாண்டு வரை ஆட்சியிலிருந்து வந்தாலும், கி.

பி. 6 ஆம் நூற்றாண்டு தொடக்கம் 9 ஆம் நூற்றாண்டு வரையான 3 நூற்றாண்டுக் காலமே காஞ்சியைத் தலைநகரமாகக் கொண்டு ஒரு வலுவுள்ள அரசை நிறுவியிருந்தனர்.

இவர்கள் தமிழர் அல்ல என்பது பொதுவாக ஏற்றுக்கொள்ளப்படினும், இவர்கள் பூர்வீகம் பற்றியோ, தமிழ் நாட்டுக்கு இவர்கள் வந்து சேர்ந்த விதம் பற்றியோ சரியான தகவல்கள் இல்லை.

பல்லவர்கள் காலத்தை முற்காலப் பல்லவர் காலம், இடைக் காலப் பல்லவர் காலம், பிற்காலப் பல்லவர் காலம் என மூன்று பிரிவுகளாகப் பிரிக்கப்படுவது உண்டு. கி. பி. 7 ஆம் நூற்றாண்டின் துவக்கத்தில் மகேந்திரவர்மனின் ஆட்சியுடன் பிற்காலப் பல்லவர் காலம் ஆரம்பமானது.

பல்லவர்களின் பொற்காலத்தின் ஆரம்பமும் அதுவே.

பிற்காலப் பல்லவர்கள் காலத்திலேயே தமிழ் நாட்டில் இந்து சமய மறுமலர்ச்சியும், சிற்பம், ஓவியம், கட்டிடக்கலை முதலிய கலை வளர்ச்சியும் உச்ச நிலையடைந்தன.

மகேந்திர பல்லவனுடன் துவங்கிய பிற்காலப் பல்லவ மன்னர் வரிசையிலே, மாமல்லன் எனப்பட்ட முதலாம் நரசிம்மன், இராஜசிம்மன் எனப்பட்ட இரண்டாம் நரசிம்மன், நந்திவர்மன் என்பவர்கள் குறிப்பிடத் தக்கவர்கள் ஆவர்.

பல்லவர் கட்டிடக்கலையை இரண்டு கட்டங்களாகப் பிரிப்பதுண்டு.

முதல் கட்டம் சுமார் கி. பி. 610 தொடக்கம் கி.

பி. 690 வரையான காலமாகும்.

இக்கட்டத்தில் அமைக்கப்பட்டவை அனைத்துமே குடைவரை கோயில்கள் ஆகும்.

இரண்டாம் கட்டமான கி. பி. 690 இலிருந்து கி.

பி. 900 வரையான காலப்பகுதியில் கட்டப்பட்டவை அனைத்தும் கட்டுமானக் (Structural) கட்டிடங்களாகும்.

இவற்றுள் முதல் கட்டத்தில் கட்டப்பட்டவை மகேந்திரன் கட்டிடங்கள், மாமல்லன் கட்டிடங்கள் இரண்டு பிரிவுகளாகப் பிரிக்கப்படுகின்றன.

முதல் பிரிவில் தூண்கள் அமைந்த மண்டபங்கள் மட்டுமே அடங்குகின்றன.

இரண்டாம் பிரிவில் மண்டபங்களுடன் இரதங்கள் எனப்படும் ஒற்றைக்கல் கோயில்களும் அடக்கம்.

இரண்டாம் கட்டமும், நரசிம்மன் கட்டிடங்கள், நந்திவர்மன் கட்டிடங்கள் என இரண்டு பிரிவுகளாக வகைப்படுத்தப்படுகின்றன.

இவ்விரண்டிலும் அடங்குபவை கட்டுமானக் கோயில்களாகும்.

பழங்காலத்தில் தமிழ் நாட்டில் கட்டிடங்கள் அழியக்கூடிய கட்டிடப்பொருட்களான மரம், சதை மற்றும் செங்கல்களினால் அமைக்கப்பட்டதாகப் பண்டைய தமிழ் இலக்கியங்களில் குறிப்புகள் காணப்படுகின்றன.

வடக்கு மற்றும் மேற்கு இந்தியப் பகுதிகளில் அசோகப் பேரரசன் காலம் முதலும் பின்னர் ஆந்திரப் பகுதிகளிலும் பயன்பாட்டிலிருந்து வந்த பெரிய பாறைகளைக் குடைந்து குடைவரைகளை அமைக்கும் முறை தமிழ் நாட்டில் கிடையாது.

பி ஆறாம் நூற்றாண்டு வரை பயன்பாட்டுக்கு வரவில்லை.

இதற்குப் பல பண்பாட்டு, பொருளாதார மற்றும் தொழில் நுட்பக் காரணிகளை ஆய்வாளர்கள் எடுத்துக் காட்டுகிறார்கள்.

ஆறாம் நூற்றாண்டளவில் தென் தமிழ் நாட்டில் பாண்டியர்களினால் குடைவரைகள் அறிமுகப் படுத்தப்பட்டதாகக் கருதப்படினும், தமிழ் நாட்டில், தேவையான சூழலும், வளங்களும், அரசியல் ஆதரவும் அளித்து இதன் வளர்ச்சிக்குப் பெருமளவு ஊக்கம் கொடுத்த பெருமை பல்லவர்களையே சாரும்.

தமிழ் நாட்டுக்கு வடக்கே ஏற்கனவே முகம் காட்டத் தொடங்கியிருந்த திராவிடக் கட்டிடக்கலை என இன்று குறிக்கப்படும் புதிய தென்னிந்தியக் கட்டிடக் கலைப் பாணிக்குத் தமிழ் நாட்டில் உரமளித்து, இந்தியாவின் முக்கியமான கட்டிடக் கலைப் பாணிகளில் ஒன்றாக வளர்வதற்கு அடித்தளமிட்டதும் பல்லவர் காலமேயாகும்.

பல்லவர்கள் அமைத்த குடைவரைகள் தமிழ் நாட்டில், மண்டகப்பட்டு, பல்லாவரம், மாமண்டூர், குரங்கணில் முட்டம், வல்லம், மகேந்திரவாடி, தளவானூர், திருச்சிராப்பள்ளி, சீயமங்கலம், விளாப்பாக்கம், அரகண்டநல்லூர், திருக்கழுக்குன்றம், சிங்கப்பெருமாள் கோயில், சிங்கவரம், மேலச்சேரி, சாமுவன் குப்பம், கீழ்மாவிலங்கை, மாமல்லபுரம், ஆலூர், திரைக்கோயில், புதூர் ஆகிய இடங்களில் காணப்படுகின்றன.

பஞ்சபாண்டவர் ரதங்கள் - மாமல்லபுரம்

தம்புள்ளையில் அமைந்துள்ள நாலந்த சிலை மண்டபம் இலங்கை கட்டிடக்கலையில் பல்லவ கட்டிடக்கலையை உள்வாங்கி அமைக்கப்பட்டுள்ளது.

Word Tokenization

In [10]:

```
1 from indicnlp.tokenize import indic_tokenize
2 from indicnlp.tokenize import indic_detokenize
3
4 string=data
5
6 print('\nInput String: {}'.format(string))
7 print('\nTokens: ')
8 for t in indic_tokenize.trivial_tokenize(string):
9     print(t)
```

Input String:

தென்னிந்தியக் கட்டிடக்கலையில் ஒரு திருப்புமுனையாக அமைந்து அதன் வளர்ச்சியில் ஆகக் கூடிய தாக்கத்தை ஏற்படுத்தியது பல்லவர் காலமே எனில் அது மிகையல்ல. இப்பகுதியில் கி.பி. 6 ஆம் நூற்றாண்டளவில் ஆரம்பித்து வளர்ச்சியடைந்த திராவிடக் கட்டிடக்கலையின் ஆரம்பம் இந்தக் காலமே.

பல்லவ வம்சத்தினர் தமிழகத்தின் வடபகுதியில் கி.பி. 4 ஆம் நூற்றாண்டு தொடக்கம் கி.பி. 10 ஆம் நூற்றாண்டு வரை ஆட்சியிலிருந்து வந்தாலும், கி.பி. 6 ஆம் நூற்றாண்டு தொடக்கம் 9 ஆம் நூற்றாண்டு வரையான 3 நூற்றாண்டுக் காலமே காஞ்சியைத் தலைநகரமாகக் கொண்டு ஒரு வலுவுள்ள அரசை நிறுவியிருந்தனர். இவர்கள் தமிழர் அல்ல என்பது பொதுவாக ஏற்றுக்கொள்ளப்படினும், இவர்கள் பூர்வீகம் பற்றியோ, தமிழ் நாட்டுக்கு இவர்கள் வந்து சேர்ந்த விதம் பற்றியோ சரியான தகவல்கள் இல்லை. பல்லவர்கள் காலத்தை முற்காலப் பல்லவர் காலம், இடைக் காலப் பல்லவர் காலம், பிற்காலப் பல்லவர் காலம் என மூன்று பிரிவுகளாகப் பிரிக்கப்படுவது உண்டு. கி.பி. 7 ஆம் நூற்றாண்டின் துவக்கத்தில் மகேந்திரவர்மனின் ஆட்சியுடன் பிற்காலப் பல்லவர் காலம் வலுவானது. பல்லவர்களின் பொற்காலத்தின் வளர்ச்சி...

Syllabification

In [11]:

```
1 from indicnlp.syllable import syllabifier
2
3 text ='சிட்டுக்குருவி மனித குடியிருப்புடன் வலுவாகத் தொடர்புடையது ஆகுமீ'
4 lang='ta'
5
6 syl_word = (' '.join(syllabifier.orthographic_syllabify(text,lang)))
7 print("Syllabification (Approximation)\n")
8 print(syl_word)## Transliteration
```

Syllabification (Approximation)

சி ட்டு க் கு ரு வி ம னி த கு டி யி ரு ப் பு ட ன் வ லு வாக த் தொ
ட ர்பு டை ய து ஆ கு மீ

Morphological Analyser

In [12]:

```
1 from indicnlp.morph import unsupervised_morph
2 from indicnlp import common
3
4 analyzer=unsupervised_morph.UnsupervisedMorphAnalyzer('ta')
```

In [13]:

```
1 string = 'புத்துணர்ச்சியான'
2
3 analyzes_tokens = analyzer.morph_analyze_document(string.split(' '))
4 print(analyzes_tokens)
```

['புத்துணர்ச்சி', 'யான']

In [14]:

```
1 text ='சிட்டுக்குருவி மனித குடியிருப்புடன் வலுவாகத் தொடர்புடையது ஆகும்'
2
3 analyzes_tokens=analyzer.morph_analyze_document(text.split(' '))
4
5 for w in analyzes_tokens:
6     print(w)
```

சிட்டுக்குருவி
மனித
குடியிருப்பு
டன்
வலுவாக
த்
தொடர்புடைய
து
ஆகும்
ி

POS Tagging - தமிழ்

In [15]:

```
1 from rippletagger.tagger import Tagger
2
3 def pos_tag_tamil(query):
4     tagger = Tagger(language="tam")
5     posTagger = tagger.tag(query)
6     return(posTagger)
```

In [16]:

```
1 query ='சிட்டுக்குருவி மனித குடியிருப்புடன் வலுவாகத் தொடர்புடையது ஆகும்'
2 tamil_tag = pos_tag_tamil(query)
3 print(tamil_tag)
```

[('சிட்டுக்குருவி', 'NOUN'), ('மனித', 'PROPN'), ('குடியிருப்புடன்', 'NOUN'), ('வலுவாகத்', 'PROPN'), ('தொடர்புடையது', 'VERB'), ('ஆகும்', 'VERB')]

In [17]:

```
1 query = "தமிழ் இலக்கியம் இரண்டாயிரம் ஆண்டுகளுக்கு மேலான தொடர்ச்சி கொண்  
2 tamil_tag = pos_tag_tamil(query)  
3 print(tamil_tag)
```

```
[('தமிழ்', 'PROPN'), ('இலக்கியம்', 'NOUN'), ('இரண்டாயிரம்', 'NOUN'),  
( 'ஆண்டுகளுக்கு', 'NOUN'), ('மேலான', 'ADJ'), ('தொடர்ச்சி', 'NOUN'),  
( 'கொண்ட', 'ADP'), ('உலகின்', 'NOUN'), ('சிறந்த', 'ADJ'), ('இலக்கிய  
ங்களில்', 'NOUN'), ('ஒன்று.', 'NOUN'), ('வாழ்வின்', 'NOUN'), ('பல்வேறு',  
'ADJ'), ('கூறுகளை', 'NOUN'), ('தமிழ்', 'PROPN'), ('இலக்கியங்கள்', 'NOUN'),  
( 'இயம்புகின்றன.', 'NOUN'), ('தமிழ்', 'PROPN'), ('இலக்கியத்தில்',  
'NOUN'), ('வெண்பா,', 'NOUN'), ('குறள்,', 'NOUN'), ('புதுக்கவிதை,', 'NOUN'),  
( 'கட்டுரை,', 'NOUN'), ('பழமொழி,', 'NOUN'), ('தொண்ணூற்றாறு',  
'NOUN'), ('வகை', 'NOUN'), ('சிறுநிலக்கியங்கள்', 'NOUN'), ('என', 'PART'),  
( 'பல', 'ADJ'), ('வடிவங்கள்', 'NOUN'), ('உள்ளன.', 'NOUN'), ('த  
மிழில்', 'NOUN'), ('வாய்மொழி', 'NOUN'), ('இலக்கியங்களும்', 'ADJ'),  
( 'முக்கிய', 'ADJ'), ('இடம்', 'ADP'), ('வகிக்கின்றன.', 'VERB')]
```

In [18]:

```
1 query = ""  
2 தென்னிந்தியக் கட்டிடக்கலையில் ஒரு திருப்புமுனையாக அமைந்து அதன் வளர்ச்சியி  
3  
4 பல்லவ வம்சத்தினர் தமிழகத்தின் வடபகுதியில் கி.பி. 4 ஆம் நூற்றாண்டு தொடக்கம்  
5  
6 பிற்காலப் பல்லவர்கள் காலத்திலேயே தமிழ் நாட்டில் இந்து சமய மறுமலர்ச்சியும், சி  
7  
8 பல்லவர் கட்டிடக்கலையை இரண்டு கட்டங்களாகப் பிரிப்பதுண்டு. முதல் கட்டம் சும  
9  
10 இவற்றுள் முதல் கட்டத்தில் கட்டப்பட்டவை மகேந்திரன் கட்டிடங்கள், மாமல்லன் கட்ட  
11  
12 இரண்டாம் கட்டமும், நரசிம்மன் கட்டிடங்கள், நந்திவர்மன் கட்டிடங்கள் என இரண்டு  
13  
14 பழங்காலத்தில் தமிழ் நாட்டில் கட்டிடங்கள் அழியக்கூடிய கட்டிடப்பொருட்களான மர  
15  
16 தமிழ் நாட்டுக்கு வடக்கே ஏற்கனவே முகம் காட்டத் தொடங்கியிருந்த திராவிடக் கட்டி  
17  
18  
19 பல்லவர்கள் அமைத்த குடைவரைகள் தமிழ் நாட்டில், மண்டகப்பட்டு, பல்லாவரம், ம  
20  
21 பஞ்சபாண்டவர் ரதங்கள் - மாமல்லபுரம்  
22  
23  
24 தம்புள்ளையில் அமைந்துள்ள நாலந்த சிலை மண்டபம் இலங்கை கட்டிடக்கலையில்  
25  
26  
27  
28 ""
```

In [19]:

```
1 tamil_tag = pos_tag_tamil(query)
2 print(tamil_tag)
```

[('தென்னிந்தியக்', 'NOUN'), ('கட்டிடக்கலையில்', 'NOUN'), ('ஒரு', 'ADJ'), ('திருப்புமுனையாக', 'ADV'), ('அமைந்து', '_'), ('அதன்', '_'), ('வளர்ச்சியில்', 'NOUN'), ('ஆகக்', 'PART'), ('கூடிய', 'ADJ'), ('தாக்கத்தை', 'NOUN'), ('ஏற்படுத்தியது', '_'), ('பல்லவர்', 'NOUN'), ('காலமே', 'NOUN'), ('எனில்', 'NOUN'), ('அது', 'PRON'), ('மிகையல்ல.', 'NOUN'), ('இப்பகுதியில்', 'NOUN'), ('கி.பி.', 'PROPN'), ('6', 'NUM'), ('ஆம்', 'AUX'), ('நூற்றாண்டளவில்', 'NOUN'), ('ஆரம்பித்து', '_'), ('வளர்ச்சியடைந்த', 'ADJ'), ('திராவிடக்', 'NOUN'), ('கட்டிடக்கலையின்', 'NOUN'), ('ஆரம்பம்', 'NOUN'), ('இந்தக்', 'NOUN'), ('காலமே.', 'NOUN'), ('பல்லவ', 'NOUN'), ('வம்சத்தினர்', 'NOUN'), ('தமிழகத்தின்', 'PROPN'), ('வடபகுதியில்', 'NOUN'), ('கி.பி.', 'PROPN'), ('4', 'NUM'), ('ஆம்', 'AUX'), ('நூற்றாண்டு', 'NOUN'), ('தொடக்கம்', 'NOUN'), ('கி.பி.', 'PROPN'), ('10', 'NUM'), ('ஆம்', 'AUX'), ('நூற்றாண்டு', 'NOUN'), ('வரை', 'ADP'), ('ஆட்சியிலிருந்து', '_'), ('வந்தாலும்', 'NOUN'), ('கி.பி.', 'PROPN'), ('6', 'NUM'), ('ஆம்', 'AUX'), ('நூற்றாண்டு', 'NOUN'), ('தொடக்கம்', 'NOUN'), ('9', 'NUM'), ('ஆம்', 'AUX'), ('நூற்றாண்டு', 'NOUN'), ('வரையான', 'ADJ'), ('3', 'NUM'), ('நூற்றாண்டுக்', 'NOUN'), ('காலமே', 'NOUN'), ('காஞ்சியைத்', 'NOUN'), ('தலைநகரமாகக்', 'ADV'), ('கொண்டு', 'VERB'), ('ஒரு', 'ADJ'), ('வலுவுள்ள', '_'), ('அரசை', 'NOUN'), ('நிறுவியிருந்தனர்.', 'PROPN'), ('இவர்கள்', 'NOUN'), ('தமிழர்', 'PROPN'), ('அல்ல', 'AUX'), ('என்பது', 'PART'), ('பொதுவாக', 'ADV'), ('ஏற்றுக்கொள்ளப்படினும்', 'NOUN'), ('இவர்கள்', 'NOUN'), ('பூர்வீகம்', 'NOUN'), ('பற்றியோ', 'NOUN'), ('தமிழ்', 'PROPN'), ('நாட்டுக்கு', 'NOUN'), ('இவர்கள்', 'NOUN'), ('வந்து', 'AUX'), ('சேர்ந்த', 'ADJ'), ('விதம்', 'NOUN'), ('பற்றியோ', 'NOUN'), ('சரியான', 'ADJ'), ('தகவல்கள்', 'NOUN'), ('இல்லை.', 'NOUN'), ('பல்லவர்கள்', 'NOUN'), ('காலத்தை', 'NOUN'), ('முற்காலப்', 'NOUN'), ('பல்லவர்', 'NOUN'), ('காலம்', 'NOUN'), ('இடைக்', 'NOUN'), ('காலப்', 'NOUN'), ('பல்லவர்', 'NOUN'), ('காலம்', 'NOUN'), ('பிற்காலப்', 'NOUN'), ('பல்லவர்', 'NOUN'), ('காலம்', 'NOUN'), ('என', 'PART'), ('மூன்று', 'NUM'), ('பிரிவுகளாகப்', 'NOUN'), ('பிரிக்கப்படுவது', '_'), ('உண்டு.', 'NOUN'), ('கி.பி.', 'PROPN'), ('7', 'NUM'), ('ஆம்', 'AUX'), ('நூற்றாண்டின்', 'NOUN'), ('துவக்கத்தில்', 'NOUN'), ('மகேந்திரவர்மனின்', 'NOUN'), ('ஆட்சியுடன்', 'NOUN'), ('பிற்காலப்', 'NOUN'), ('பல்லவர்', 'NOUN'), ('காலம்', 'NOUN'), ('ஆரம்பமானது.', 'NOUN'), ('பல்லவர்களின்', 'NOUN'), ('பொற்காலத்தின்', 'NOUN'), ('ஆரம்பமும்', '_'), ('அதுவே.', 'NOUN'), ('பிற்காலப்', 'NOUN'), ('பல்லவர்கள்', 'NOUN'), ('காலத்திலேயே', 'NOUN'), ('தமிழ்', 'PROPN'), ('நாட்டில்', 'NOUN'), ('இந்து', 'PROPN'), ('சமய', 'NOUN'), ('மறுமலர்ச்சியும்', 'NOUN'), ('சிற்பம்', 'NOUN'), ('ஓவியம்', 'NOUN'), ('கட்டிடக்கலை', 'NOUN'), ('முதலிய', 'ADJ'), ('கலை', 'NOUN'), ('வளர்ச்சியும்', 'ADJ'), ('உச்ச', 'PROPN'), ('நிலையடைந்தன.', 'NOUN'), ('மகேந்திர', 'PROPN'), ('பல்லவனுடன்', 'NOUN'), ('துவங்கிய', 'ADJ'), ('பிற்காலப்', 'NOUN'), ('பல்லவ', 'NOUN'), ('மன்னர்', 'NOUN'), ('வரிசையிலே', 'NOUN'), ('மாமல்லன்', 'NOUN'), ('எனப்பட்ட', 'ADJ'), ('முதலாம்', 'NOUN'), ('நரசிம்மன்', 'NOUN'), ('இராஜசிம்மன்', 'NOUN'), ('எனப்பட்ட', 'ADJ'), ('இரண்டாம்', 'ADJ'), ('நரசிம்மன்', 'NOUN'), ('நந்திவர்மன்', 'NOUN'), ('என்பவர்கள்', 'NOUN'), ('குறிப்பிடத்', 'VERB'), ('தக்கவர்கள்', 'NOUN'), ('ஆவர்.', 'PROPN'), ('பல்லவர்', 'NOUN'), ('கட்டிடக்கலையை', 'NOUN'), ('இரண்டு', 'NUM'), ('கட்டங்களாகப்', 'NOUN'), ('பிரிப்பதுண்டு.', 'NOUN'), ('முதல்', 'ADJ'), ('கட்டம்', 'NOUN'), ('சுமார்', 'ADV'), ('கி.பி.', 'PROPN'), ('610', 'NUM'), ('தொடக்கம்', 'NOUN'), ('கி.பி.', 'PROPN'), ('690', 'NUM'), ('வரையான', 'ADJ'), ('காலமாகும்.', 'PROPN'), ('இக்கட்டத்தில்', 'NOUN'), ('அமைக்கப்பட்டவை', 'NOUN'), ('அனைத்துமே', '_'), ('குடைவரை', 'NOUN'), ('கோயில்கள்', 'NOUN'), ('ஆகும்.', 'PROPN'), ('இரண்டாம்', 'ADJ'), ('கட்டமான', 'ADJ'), ('கி.பி.', 'PROPN'), ('690', 'NUM'), ('இலிருந்து', 'ADP'), ('கி.பி.', 'PROPN'), ('900', 'NUM'), ('வரையான', 'ADJ'), ('காலப்பகுதியில்', 'NOUN'),

('கட்டப்பட்டவை', 'NOUN'), ('அனைத்தும்', 'PRON'), ('கட்டுமானக்', 'NOUN'), ('(Structural)', 'NOUN'), ('கட்டிடங்களாகும்.', 'PROPN'), ('இவற்றுள்', 'NOUN'), ('முதல்', 'ADJ'), ('கட்டத்தில்', 'NOUN'), ('கட்டப்பட்டவை', 'NOUN'), ('மகேந்திரன்', 'NOUN'), ('கட்டிடங்கள்', 'NOUN'), ('மாமல்லன்', 'NOUN'), ('கட்டிடங்கள்', 'NOUN'), ('இரண்டு', 'NUM'), ('பிரிவுகளாகப்', 'NOUN'), ('பிரிக்கப்படுகின்றன.', 'NOUN'), ('முதல்', 'ADJ'), ('பிரிவில்', 'NOUN'), ('தூண்கள்', 'NOUN'), ('அமைந்த', '_'), ('மண்டபங்கள்', 'NOUN'), ('மட்டுமே', 'PART'), ('அடங்குகின்றன.', 'NOUN'), ('இரண்டாம்', 'ADJ'), ('பிரிவில்', 'NOUN'), ('மண்டபங்களுடன்', 'NOUN'), ('இரதங்கள்', 'NOUN'), ('எனப்படும்', 'ADJ'), ('ஒற்றைக்கல்', 'NOUN'), ('கோயில்களும்', 'ADV'), ('அடக்கம்.', 'PROPN'), ('இரண்டாம்', 'ADJ'), ('கட்டமும்', 'NOUN'), ('நரசிம்மன்', 'NOUN'), ('கட்டிடங்கள்', 'NOUN'), ('நந்திவர்மன்', 'NOUN'), ('கட்டிடங்கள்', 'NOUN'), ('என', 'PART'), ('இரண்டு', 'NUM'), ('பிரிவுகளாக', 'ADV'), ('வகைப்படுத்தப்படுகின்றன.', 'NOUN'), ('இவ்விரண்டிலும்', 'ADJ'), ('அடங்குபவை', 'NOUN'), ('கட்டுமானக்', 'NOUN'), ('கோயில்களாகும்.', 'PROPN'), ('பழங்காலத்தில்', 'NOUN'), ('தமிழ்', 'PROPN'), ('நாட்டில்', 'NOUN'), ('கட்டிடங்கள்', 'NOUN'), ('அழியக்கூடிய', 'ADJ'), ('கட்டிடப்பொருட்களான', 'ADJ'), ('மரம்', 'NOUN'), ('சுதை', 'NOUN'), ('மற்றும்', 'CONJ'), ('செங்கல்களினால்', 'NOUN'), ('அமைக்கப்பட்டதாகப்', 'NOUN'), ('பண்டைய', 'NOUN'), ('தமிழ்', 'PROPN'), ('இலக்கியங்களில்', 'NOUN'), ('குறிப்புகள்', 'NOUN'), ('காணப்படுகின்றன.', 'NOUN'), ('வடக்கு', 'NOUN'), ('மற்றும்', 'CONJ'), ('மேற்கு', 'VERB'), ('இந்தியப்', 'PROPN'), ('பகுதிகளில்', 'NOUN'), ('அசோகப்', 'NOUN'), ('பேரரசன்', 'NOUN'), ('காலம்', 'NOUN'), ('முதலும்', 'ADJ'), ('பின்னர்', 'ADV'), ('ஆந்திரப்', 'NOUN'), ('பகுதிகளிலும்', '_'), ('பயன்பாட்டிலிருந்து', '_'), ('வந்த', 'ADJ'), ('பெரிய', 'ADJ'), ('பாறைகளைக்', 'NOUN'), ('குடைந்து', '_'), ('குடைவரைகளை', 'NOUN'), ('அமைக்கும்', 'ADJ'), ('முறை', 'NOUN'), ('தமிழ்', 'PROPN'), ('நாட்டில்', 'NOUN'), ('கி.பி', 'NOUN'), ('ஆறாம்', 'NOUN'), ('நூற்றாண்டு', 'NOUN'), ('வரை', 'ADP'), ('பயன்பாட்டுக்கு', 'NOUN'), ('வரவில்லை.', 'NOUN'), ('இதற்குப்', 'NOUN'), ('பல', 'ADJ'), ('பண்பாட்டு', 'NOUN'), ('பொருளாதார', 'PROPN'), ('மற்றும்', 'CONJ'), ('தொழில்', 'NOUN'), ('நுட்பக்', 'NOUN'), ('காரணிகளை', 'NOUN'), ('ஆய்வாளர்கள்', 'NOUN'), ('எடுத்துக்', 'NOUN'), ('காட்டுகிறார்கள்.', 'PROPN'), ('ஆறாம்', 'NOUN'), ('நூற்றாண்டளவில்', 'NOUN'), ('தென்', 'NOUN'), ('தமிழ்', 'PROPN'), ('நாட்டில்', 'NOUN'), ('பாண்டியர்களினால்', 'NOUN'), ('குடைவரைகள்', 'NOUN'), ('அறிமுகப்', 'NOUN'), ('படுத்தப்பட்டதாகக்', 'ADV'), ('கருதப்பட்டனும்', 'NOUN'), ('தமிழ்', 'PROPN'), ('நாட்டில்', 'NOUN'), ('தேவையான', 'ADJ'), ('சூழலும்', 'NOUN'), ('வளங்களும்', 'NOUN'), ('அரசியல்', 'NOUN'), ('ஆதரவும்', 'PRON'), ('அளித்து', '_'), ('இதன்', 'PROPN'), ('வளர்ச்சிக்குப்', 'NOUN'), ('பெருமளவு', 'NOUN'), ('ஊக்கம்', 'NOUN'), ('கொடுத்த', 'ADJ'), ('பெருமை', 'NOUN'), ('பல்லவர்களையே', 'NOUN'), ('சாரும்.', 'PROPN'), ('தமிழ்', 'PROPN'), ('நாட்டுக்கு', 'NOUN'), ('வடக்கே', 'NOUN'), ('ஏற்கனவே', 'ADV'), ('முகம்', 'NOUN'), ('காட்டத்', 'VERB'), ('தொடங்கியிருந்த', 'ADJ'), ('திராவிடக்', 'NOUN'), ('கட்டிடக்கலை', 'NOUN'), ('என', 'PART'), ('இன்று', 'ADV'), ('குறிக்கப்படும்', 'ADJ'), ('புதிய', 'ADJ'), ('தென்னிந்தியக்', 'NOUN'), ('கட்டிடக்', 'NOUN'), ('கலைப்', 'NOUN'), ('பாணிக்குத்', 'NOUN'), ('தமிழ்', 'PROPN'), ('நாட்டில்', 'NOUN'), ('உரமளித்து', 'NOUN'), ('இந்தியாவின்', 'PROPN'), ('முகக்கியமான', 'ADJ'), ('கட்டிடக்', 'NOUN'), ('கலைப்', 'NOUN'), ('பாணிகளில்', 'NOUN'), ('ஒன்றாக', 'ADV'), ('வளர்வதற்கு', 'VERB'), ('அடித்தளமிட்டதும்', 'ADJ'), ('பல்லவர்', 'NOUN'), ('காலமேயாகும்.', 'PROPN'), ('பல்லவர்கள்', 'NOUN'), ('அமைத்த', 'ADJ'), ('குடைவரைகள்', 'NOUN'), ('தமிழ்', 'PROPN'), ('நாட்டில்', 'NOUN'), ('மண்டகப்பட்டு', 'NOUN'), ('பல்லாவரம்', 'NOUN'), ('மாமண்டூர்', 'NOUN'), ('குரங்கணில்', 'NOUN'), ('முட்டம்', 'NOUN'), ('வல்லம்', 'NOUN'), ('மகேந்திரவாடி', 'NOUN'), ('தளவானூர்', 'NOUN'), ('திருச்சிராப்பள்ளி', 'NOUN'), ('சீயமங்கலம்', 'NOUN'), ('விளாப்பாக்கம்', 'NOUN'), ('அரகண்டநல்லூர்', 'NOUN'), ('திருக்கழுக்குன்றம்', 'NOUN'), ('சிங்கப்பெருமாள்', 'NOUN'), ('கோயில்', 'NOUN'), ('சிங்கவரம்', 'NOUN'), ('மேலச்சேரி', 'NOUN'), ('சாமுவன்', 'NOUN'), ('குப்பம்', 'NOUN'), ('கீழ்மாவிலங்கை', 'NOUN'), ('மாமல்லபுரம்', 'NOUN'),

```
('NOUN'), ('ஆவூர்,', 'NOUN'), ('திரைக்கோயில்,', 'NOUN'), ('புதூர்', 'PROP  
N'), ('ஆகிய', 'ADJ'), ('இடங்களில்', 'NOUN'), ('காணப்படுகின்றன.', 'N  
OUN'), ('பஞ்சபாண்டவர்', 'NOUN'), ('ரதங்கள்', 'NOUN'), ('-', 'PUNCT'),  
('மாமல்லபுரம்', 'NOUN'), ('தம்புள்ளையில்', 'NOUN'), ('அமைந்துள்ள',  
'_'), ('நாலந்த', 'ADJ'), ('சிலை', 'NOUN'), ('மண்டபம்', 'NOUN'), ('இல  
ங்கை', 'PROP'), ('கட்டிடக்கலையில்', 'NOUN'), ('பல்லவ', 'NOUN'), ('க  
ட்டிடக்கலையை', 'NOUN'), ('உள்வாங்கி', 'VERB'), ('அமைக்கப்பட்டுள்ள  
து.', 'VERB')]
```

StopWords - தமிழ்

In [20]:

```
1 with open('TamilStopWords.txt', encoding="utf8") as file:  
2     stop_words_tamil = file.read()  
3  
4 print(stop_words_tamil)
```

ஒரு
என்று
மற்றும்
இந்த
இது
என்ற
கொண்டு
என்பது
பல
ஆகும்
அல்லது
அவர்
நான்
உள்ள
அந்த
இவர்
என
முதல்
என்ன
கொண்ட

In [21]:

```
1 data="""தமிழ் இலக்கியம் இரண்டாயிரம் ஆண்டுகளுக்கு மேலான தொடர்ச்சி கொண்  
2
```

In [22]:

```
1 token = []  
2 for t in indic_tokenize.trivial_tokenize(data):  
3     token.append(t)
```

In [23]:

```
1 print("Tokenized Words :\n ")
2 print(token)
```

Tokenized Words :

```
['"', 'தமிழ்', 'இலக்கியம்', 'இரண்டாயிரம்', 'ஆண்டுகளுக்கு', 'மே  
லான', 'தொடர்ச்சி', 'கொண்ட', 'உலகின்', 'சிறந்த', 'இலக்கியங்க  
ளில்', 'ஒன்று', '.', 'வாழ்வின்', 'பல்வேறு', 'கூறுகளை', 'தமிழ்', 'இல  
க்கியங்கள்', 'இயம்புகின்றன', '.', 'தமிழ்', 'இலக்கியத்தில்', 'வெ  
ண்பா', ',', 'குறள்', ',', 'புதுக்கவிதை', ',', 'கட்டுரை', ',', 'பழ  
மொழி', ',', 'தொண்ணூற்றாறு', 'வகை', 'சிற்றிலக்கியங்கள்', 'என',  
'பல', 'வடிவங்கள்', 'உள்ளன', '.', 'தமிழில்', 'வாய்மொழி', 'இலக்கிய  
ங்களும்', 'முக்கிய', 'இடம்', 'வகிக்கின்றன', '.']
```

In [24]:

```
1 fil_sen = []
2 fil_sw = []
3
4 for t in indic_tokenize.trivial_tokenize(data):
5     if t not in stop_words_tamil:
6         fil_sen.append(t)
7     else:
8         fil_sw.append(t)
```

In [25]:

```
1 print(fil_sen)
```

```
['"', 'தமிழ்', 'இலக்கியம்', 'இரண்டாயிரம்', 'ஆண்டுகளுக்கு', 'மே  
லான', 'தொடர்ச்சி', 'உலகின்', 'சிறந்த', 'இலக்கியங்களில்', 'ஒன்று',  
, '.', 'வாழ்வின்', 'கூறுகளை', 'தமிழ்', 'இலக்கியங்கள்', 'இயம்புகின்ற  
ன', '.', 'தமிழ்', 'இலக்கியத்தில்', 'வெண்பா', ',', 'குறள்', ',', 'பு  
துக்கவிதை', ',', 'கட்டுரை', ',', 'பழமொழி', ',', 'தொண்ணூற்றாறு',  
'வகை', 'சிற்றிலக்கியங்கள்', 'வடிவங்கள்', '.', 'தமிழில்', 'வா  
ய்மொழி', 'இலக்கியங்களும்', 'முக்கிய', 'வகிக்கின்றன', '.']
```

In [26]:

```
1 print("Removed Stop Words :\n ")
2 print(fil_sw)
```

Removed Stop Words :

```
['கொண்ட', 'பல்வேறு', 'என', 'பல', 'உள்ளன', 'இடம்']
```

In [27]:

```
1 print("Number of Words including StopWords : ", len(token))
2 print("Number of Words excluding StopWords : ", len(fil_sen))
3 print("Total number of Stop Words Excluded : ", len(fil_sw))
```

```
Number of Words including StopWords : 47
Number of Words excluding StopWords : 41
Total number of Stop Words Excluded : 6
```

Transliteration

In [28]:

```
1 from indicnlp.transliterate.unicode_transliterate import UnicodeIndicTransliterator
2
3 input_text = 'சிட்டுக்குருவி மனித குடியிருப்புடன் வலுவாகத் தொடர்புடையது ஆகும்'
4 ta_hi_text = (UnicodeIndicTransliterator.transliterate(input_text,"ta","hi"))
5
6 print("Transliterate From Tamil to Hindi\n")
7 print("Input Text : " , input_text)
8 print("output Text : ", ta_hi_text)
```

Transliterate From Tamil to Hindi

Input Text : சிட்டுக்குருவி மனித குடியிருப்புடன் வலுவாகத் தொடர்புடையது ஆகும்
output Text : चिट्ठुकुरुवि मनिठ कुटियिरुप्पुटन् वलुवाकत् तोटपुटैयतु आकुम्

In [29]:

```
1 from indicnlp.transliterate.unicode_transliterate import UnicodeIndicTransliterator
2
3 input_text=ta_hi_text
4 hi_ta_text = (UnicodeIndicTransliterator.transliterate(input_text,"hi","ta"))
5
6 print("Transliterate From Hindi to Tamil\n")
7 print("Input Text : " , input_text)
8 print("output Text : ", hi_ta_text)
```

Transliterate From Hindi to Tamil

Input Text : चिट्ठुकुरुवि मनिठ कुटियिरुप्पुटन् वलुवाकत् तोटपुटैयतु आकुम्
output Text : சிட்டுக்குருவி மனித குடியிருப்புடன் வலுவாகத் தொடர்புடையது ஆகும்

In [30]:

```
1 from indicnlp.transliterate.unicode_transliterate import ItransTransliterator
2
3 input_text = 'சிட்டுக்குருவி மனித குடியிருப்புடன் வலுவாகத் தொடர்புடையது ஆகும்'
4 lang='ta'
5
6 ta_eng = (ItransTransliterator.to_itrans(input_text,lang))
7
8 print("Transliterate From Hindi to Tamil\n")
9 print("Input Text : " , input_text)
10 print("output Text : ", ta_eng)
```

Transliterate From Hindi to Tamil

Input Text : சிட்டுக்குருவி மனித குடியிருப்புடன் வலுவாகத் தொடர்புடையது ஆகும்
output Text : chiTTukkuruvi ma*nita kuTiyiruppuTa*n valuvaakat t.oTarpuTaiyatu aakum

Translation Options

In [31]:

```
1 text ='சிட்டுக்குருவி மனித குடியிருப்புடன் வலுவாகத் தொடர்புடையது ஆகும்'
```

In [32]:

```
1 from textblob import TextBlob
2 txt = TextBlob(text)
3 # Can detect language
4 txt.detect_language()
```

Out[32]:

'ta'

In [33]:

```
1 # Language Translation also available in TextBlob
2 ta_eng_text = txt.translate(from_lang='ta', to ='en')
3 print(ta_eng_text)
```

The sparrow is strongly associated with human habitation

In [34]:

```
1 txt = ta_eng_text
2 # Language Translation also available in TextBlob
3 print(txt.translate(from_lang='en', to ='ta'))
```

குருவி மனித வாழ்விடத்துடன் வலுவாக தொடர்புடையது

In [35]:

```
1 text =""""தமிழ் இலக்கியம் இரண்டாயிரம் ஆண்டுகளுக்கு மேலான தொடர்ச்சி கொண்
2 from textblob import TextBlob
3 txt = TextBlob(text)
4 # Can detect language
5 txt.detect_language()
```

Out[35]:

'ta'

In [36]:

```
1 # Language Translation also available in TextBlob
2 ta_eng_text = txt.translate(from_lang='ta', to ='en')
3 print(ta_eng_text)
```

"Tamil literature is one of the best literature in the world with more than two thousand years of continuity.

In [37]:

```
1 text = "வாழ்வின் பல்வேறு கூறுகளை தமிழ் இலக்கியங்கள் இயம்புகின்றன. தமிழ் இ  
2 from textblob import TextBlob  
3 txt = TextBlob(text)  
4 # Can detect Language  
5 txt.detect_language()
```

Out[37]:

'ta'

In [38]:

```
1 # Language Translation also available in TextBlob  
2 ta_eng_text = txt.translate(from_lang='ta', to ='en')  
3 print(ta_eng_text)
```

Tamil literature encompasses many aspects of life. There are many forms in Tamil literature such as whitepaper, short, bibliography, essay, proverb, ninety-six short stories. Verbal literature also plays an important role in Tamil.