working with Lemmatization and Stemming Techniques

Stemming Technique is technique of reducing the word to its root word (stem)

Eg: Running -> run, Eating -> eat, eaten -> eat

Rule Based Stemming | Porter Stemmer

[4]:
```
from nltk.stem import PorterStemmer
ps = PorterStemmer()
print(ps.stem('running'))
```

```
run
```

Dictionary Based : Different Forms of verbs have different key but stem word remains same

[2]:
```
dictionary = {'running':'run','ran':'run','runner':'run'}
word = 'running'
print(dictionary.get(word,word))
```

```
run
```

Corpus Based Stemming Technique

It will pick up the word on the basis of the frequency

[4]:
```
forms = {'run':50,'running':30,'ran':20}
stem =  max(forms,key=forms.get)
print(stem)
```

```
run
```

Hybrid Stemming Technique : Prefined Technique with Custom

Predefined

[ ]:
```
from nltk.stem import PorterStemmer
ps = PorterStemmer()
```

Custom Technique

[9]:
```
dictionary = {'running':'run'}
word = 'running'
```

[10]:
```
print(dictionary.get(word,word))
```

```
run
```

Merging Both Predefined Algorithm + Customer Technique = Hybrid Stemming

[12]:
```
if word in dictionary:
    stem = dictionary[word]
else:
    stem = ps.stem(word)
print(stem)
```

```
run
```

Light Stemming Technique : we try to find stem word on basis of some prefix or suffix

suffix : post of the word prefix : pre of the word

[13]:
```
suffix = "ing"
```

[14]:
```
l = len(suffix)
```

[15]:
```
print(l)
```

```
3
```

[16]:
```
word = input('Enter the word:')
if word.endswith(suffix):
   stem = word[:l]
else:
```

```
    stem = word
print('Light Stemming:',stem)
```

```
Enter the word: eat
Light Stemming: eat
```

[17]:
```
word = input('Enter the word:')
if word.endswith(suffix):
   stem = word[:l]
else:
    stem = word
print('Light Stemming:',stem)
```

```
Enter the word: eating
Light Stemming: eat
```

```
#### Langauge-specific Morphological Stemming (Hindi Example)
```

running -> Hindi form -> chalna

Indicnlp used for Hindi NLP Techniques : https://github.com/anoopkunchukuttan/indic_nlp_library

[23]:
```
word = "ladkiyon" # girls
suffix = "yon"
l = len(suffix)
stem = word[:-1*l] if word.endswith(suffix) else word
print(stem)
```

```
ladki
```

Stemming Algorithms

- Porter Stemmer
- Snowball Stemmer
- LanCaster Stemmer

Example

```
flies => porter => fli => fli => fli
happily => happili => happi => happy
fishing => fish => fish => fish
better => better => better => bet
```

[28]:
```
from nltk.stem import PorterStemmer,SnowballStemmer,LancasterStemmer
```

[30]:
```
ps = PorterStemmer()
sb = SnowballStemmer('english') # Snowball stemmer takes english as langauge
lc = LancasterStemmer()
```

[33]:
```
words = ['running','flies','happily','fishing','better']
print(f" {'word':<20} {'Porter':<20} {'Snowball':<20} {'Lancaster':<20}")
print("-"*80)
```

```
 word                 Porter               Snowball             Lancaster
--------------------------------------------------------------------------------
```

[34]:
```
print(f" {'word':<20} {'Porter':<20} {'Snowball':<20} {'Lancaster':<20}")
print("-"*80)
## use for loop to print the output
for w in words:
    print(f" {w:<20} {ps.stem(w):<20} {sb.stem(w):<20} {lc.stem(w):<20}")
print("-"*80)
```

```
 word                 Porter               Snowball             Lancaster
--------------------------------------------------------------------------------
 running              run                  run                  run
 flies                fli                  fli                  fli
 happily              happili              happili              happy
 fishing              fish                 fish                 fish
 better               better               better               bet
--------------------------------------------------------------------------------
```

[ ]: