

Submit_Text_Processing_Code

April 24, 2020

Program to split the paragraph into sentences

```
[1]: import nltk
from nltk.tokenize import sent_tokenize
text = "For the third time this monsoon, pounding rains paralysed Mumbai on
→Wednesday with total disruption in road and rail traffic, though flights
→were not hit severely. Mumbai turned into an extension of the sea as
→unprecedented rains deluged the country's financial capital, delaying
→flights, jamming traffic, flooding homes and having rescue teams out on the
→roads since morning. In several places, the local trains grinded to a halt,
→and long distance trains were cancelled. In the evening office hours, the
→waterlogging pushed people to look for the nearest shelter, abandoning
→thoughts of reaching home. The met office declared a red alert for the next
→24 hours, indicating that the situation might worsen. Schools and junior
→colleges in Mumbai, Thane and Konkan will remain shut on Thursday as a
→precautionary measure."
sentences = nltk.sent_tokenize(text)
for sentence in sentences:
    print(sentence)
    print()
```

For the third time this monsoon, pounding rains paralysed Mumbai on Wednesday with total disruption in road and rail traffic, though flights were not hit severely.

Mumbai turned into an extension of the sea as unprecedented rains deluged the country's financial capital, delaying flights, jamming traffic, flooding homes and having rescue teams out on the roads since morning.

In several places, the local trains grinded to a halt, and long distance trains were cancelled.

In the evening office hours, the waterlogging pushed people to look for the nearest shelter, abandoning thoughts of reaching home.

The met office declared a red alert for the next 24 hours, indicating that the situation might worsen.

Schools and junior colleges in Mumbai, Thane and Konkan will remain shut on Thursday as a precautionary measure.

Program to print list of words other than stop words

```
[2]: from nltk.corpus import stopwords
nltk.download('stopwords')
#print(stopwords.words("english"))

stop_words = set(stopwords.words("english"))
sentence = "For the third time this monsoon, pounding rains paralysed Mumbai on
→Wednesday with total disruption in road and rail traffic, though flights
→were not hit severely. Mumbai turned into an extension of the sea as
→unprecedented rains deluged the country's financial capital, delaying
→flights, jamming traffic, flooding homes and having rescue teams out on the
→roads since morning. In several places, the local trains grinded to a halt,
→and long distance trains were cancelled. In the evening office hours, the
→waterlogging pushed people to look for the nearest shelter, abandoning
→thoughts of reaching home. The met office declared a red alert for the next
→24 hours, indicating that the situation might worsen. Schools and junior
→colleges in Mumbai, Thane and Konkan will remain shut on Thursday as a
→precautionary measure."

words = nltk.word_tokenize(sentence)
without_stop_words = [word for word in words if not word in stop_words]
print(without_stop_words)
```

```
['For', 'third', 'time', 'monsoon', ',', 'pounding', 'rains', 'paralysed',
'Mumbai', 'Wednesday', 'total', 'disruption', 'road', 'rail', 'traffic', ',',
'though', 'flights', 'hit', 'severely', '.', 'Mumbai', 'turned', 'extension',
'sea', 'unprecedented', 'rains', 'deluged', 'country', "'s", 'financial',
'capital', ',', 'delaying', 'flights', ',', 'jamming', 'traffic', ',',
'flooding', 'homes', 'rescue', 'teams', 'roads', 'since', 'morning', '.', 'In',
'several', 'places', ',', 'local', 'trains', 'grinded', 'halt', ',', 'long',
'distance', 'trains', 'cancelled', '.', 'In', 'evening', 'office', 'hours', ',',
'waterlogging', 'pushed', 'people', 'look', 'nearest', 'shelter', ',',
'abandoning', 'thoughts', 'reaching', 'home', '.', 'The', 'met', 'office',
'declared', 'red', 'alert', 'next', '24', 'hours', ',', 'indicating',
'situation', 'might', 'worsen', '.', 'Schools', 'junior', 'colleges', 'Mumbai',
',', 'Thane', 'Konkan', 'remain', 'shut', 'Thursday', 'precautionary',
'measure', '.']
```

```
[nltk_data] Downloading package stopwords to
[nltk_data] /home/sanjukta/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

Program for extracting nouns in a sentence

```
[3]: import nltk
from nltk.tokenize import word_tokenize
from nltk.tag import pos_tag
from nltk.chunk import ne_chunk
doc = '''For the third time this monsoon, pounding rains paralysed Mumbai on
→Wednesday with total disruption in road and rail traffic, though flights
→were not hit severely. Mumbai turned into an extension of the sea as
→unprecedented rains deluged the country's financial capital, delaying
→flights, jamming traffic, flooding homes and having rescue teams out on the
→roads since morning. In several places, the local trains grinded to a halt,
→and long distance trains were cancelled. In the evening office hours, the
→waterlogging pushed people to look for the nearest shelter, abandoning
→thoughts of reaching home. The met office declared a red alert for the next
→24 hours, indicating that the situation might worsen. Schools and junior
→colleges in Mumbai, Thane and Konkan will remain shut on Thursday as a
→precautionary measure.'''
word_tokens = nltk.word_tokenize(doc)
print(word_tokens)
```

```
['For', 'the', 'third', 'time', 'this', 'monsoon', ',', 'pounding', 'rains',
'paralysed', 'Mumbai', 'on', 'Wednesday', 'with', 'total', 'disruption', 'in',
'road', 'and', 'rail', 'traffic', ',', 'though', 'flights', 'were', 'not',
'hit', 'severely', '.', 'Mumbai', 'turned', 'into', 'an', 'extension', 'of',
'the', 'sea', 'as', 'unprecedented', 'rains', 'deluged', 'the', 'country', "'s",
'financial', 'capital', ',', 'delaying', 'flights', ',', 'jamming', 'traffic',
',', 'flooding', 'homes', 'and', 'having', 'rescue', 'teams', 'out', 'on',
'the', 'roads', 'since', 'morning', '.', 'In', 'several', 'places', ',', 'the',
'local', 'trains', 'grinded', 'to', 'a', 'halt', ',', 'and', 'long', 'distance',
'trains', 'were', 'cancelled', '.', 'In', 'the', 'evening', 'office', 'hours',
',', 'the', 'waterlogging', 'pushed', 'people', 'to', 'look', 'for', 'the',
'nearest', 'shelter', ',', 'abandoning', 'thoughts', 'of', 'reaching', 'home',
'.', 'The', 'met', 'office', 'declared', 'a', 'red', 'alert', 'for', 'the',
'next', '24', 'hours', ',', 'indicating', 'that', 'the', 'situation', 'might',
'worsen', '.', 'Schools', 'and', 'junior', 'colleges', 'in', 'Mumbai', ',',
'Thane', 'and', 'Konkan', 'will', 'remain', 'shut', 'on', 'Thursday', 'as', 'a',
'precautionary', 'measure', '.']
```

```
[4]: postags = nltk.pos_tag(word_tokens)
print(postags)
```

```
[('For', 'IN'), ('the', 'DT'), ('third', 'JJ'), ('time', 'NN'), ('this', 'DT'),
('monsoon', 'NN'), (',', ','), ('pounding', 'VBG'), ('rains', 'NNS'),
('paralysed', 'VBN'), ('Mumbai', 'NNP'), ('on', 'IN'), ('Wednesday', 'NNP'),
('with', 'IN'), ('total', 'JJ'), ('disruption', 'NN'), ('in', 'IN'), ('road',
'NN'), ('and', 'CC'), ('rail', 'NN'), ('traffic', 'NN'), (',', ','), ('though',
'IN'), ('flights', 'NNS'), ('were', 'VBD'), ('not', 'RB'), ('hit', 'VBN'),
```

```
(('severely', 'RB'), (('.', '.'), ('Mumbai', 'NNP'), ('turned', 'VBD'), ('into', 'IN'), ('an', 'DT'), ('extension', 'NN'), ('of', 'IN'), ('the', 'DT'), ('sea', 'NN'), ('as', 'IN'), ('unprecedented', 'JJ'), ('rains', 'NNS'), ('deluged', 'VBD'), ('the', 'DT'), ('country', 'NN'), ('s', 'POS'), ('financial', 'JJ'), ('capital', 'NN'), (',', ','), ('delaying', 'VBG'), ('flights', 'NNS'), (',', ','), ('jamming', 'VBG'), ('traffic', 'NN'), (',', ','), ('flooding', 'VBG'), ('homes', 'NNS'), ('and', 'CC'), ('having', 'VBG'), ('rescue', 'NN'), ('teams', 'NNS'), ('out', 'RP'), ('on', 'IN'), ('the', 'DT'), ('roads', 'NNS'), ('since', 'IN'), ('morning', 'NN'), (('.', '.'), ('In', 'IN'), ('several', 'JJ'), ('places', 'NNS'), (',', ','), ('the', 'DT'), ('local', 'JJ'), ('trains', 'NNS'), ('grinded', 'VBD'), ('to', 'TO'), ('a', 'DT'), ('halt', 'NN'), (',', ','), ('and', 'CC'), ('long', 'JJ'), ('distance', 'NN'), ('trains', 'NNS'), ('were', 'VBD'), ('cancelled', 'VBN'), (('.', '.'), ('In', 'IN'), ('the', 'DT'), ('evening', 'NN'), ('office', 'NN'), ('hours', 'NNS'), (',', ','), ('the', 'DT'), ('waterlogging', 'NN'), ('pushed', 'VBD'), ('people', 'NNS'), ('to', 'TO'), ('look', 'VB'), ('for', 'IN'), ('the', 'DT'), ('nearest', 'JJS'), ('shelter', 'NN'), (',', ','), ('abandoning', 'VBG'), ('thoughts', 'NNS'), ('of', 'IN'), ('reaching', 'VBG'), ('home', 'NN'), (('.', '.'), ('The', 'DT'), ('met', 'JJ'), ('office', 'NN'), ('declared', 'VBD'), ('a', 'DT'), ('red', 'JJ'), ('alert', 'NN'), ('for', 'IN'), ('the', 'DT'), ('next', 'JJ'), ('24', 'CD'), ('hours', 'NNS'), (',', ','), ('indicating', 'VBG'), ('that', 'IN'), ('the', 'DT'), ('situation', 'NN'), ('might', 'MD'), ('worsen', 'VB'), (('.', '.'), ('Schools', 'NNP'), ('and', 'CC'), ('junior', 'JJ'), ('colleges', 'NNS'), ('in', 'IN'), ('Mumbai', 'NNP'), (',', ','), ('Thane', 'NNP'), ('and', 'CC'), ('Konkan', 'NNP'), ('will', 'MD'), ('remain', 'VB'), ('shut', 'NNS'), ('on', 'IN'), ('Thursday', 'NNP'), ('as', 'IN'), ('a', 'DT'), ('precautionary', 'JJ'), ('measure', 'NN'), (('.', '.'))]
```

```
[5]: nouns = ne_chunk(postags)
```

```
[6]: for x in str(nouns).split('\n'):
      if '/NN' in x:
          print(x)
```

```
time/NN
monsoon/NN
rains/NNS
(PERSON Mumbai/NNP)
Wednesday/NNP
disruption/NN
road/NN
rail/NN
traffic/NN
flights/NNS
(PERSON Mumbai/NNP)
extension/NN
sea/NN
rains/NNS
```

country/NN
capital/NN
flights/NNS
traffic/NN
homes/NNS
rescue/NN
teams/NNS
roads/NNS
morning/NN
places/NNS
trains/NNS
halt/NN
distance/NN
trains/NNS
evening/NN
office/NN
hours/NNS
waterlogging/NN
people/NNS
shelter/NN
thoughts/NNS
home/NN
office/NN
alert/NN
hours/NNS
situation/NN
(PERSON Schools/NNP)
colleges/NNS
(GPE Mumbai/NNP)
(GPE Thane/NNP)
(PERSON Konkan/NNP)
shut/NNS
Thursday/NNP
measure/NN

Program to transform all verbs in the paragraph into their base forms

```
[9]: import nltk
from nltk import WordNetLemmatizer
wl=WordNetLemmatizer()
```

```

doc="For the third time this monsoon, pounding rains paralysed Mumbai on
→Wednesday with total disruption in road and rail traffic, though flights
→were not hit severely. Mumbai turned into an extension of the sea as
→unprecedented rains deluged the country's financial capital, delaying
→flights, jamming traffic, flooding homes and having rescue teams out on the
→roads since morning. In several places, the local trains grinded to a halt,
→and long distance trains were cancelled. In the evening office hours, the
→waterlogging pushed people to look for the nearest shelter, abandoning
→thoughts of reaching home. The met office declared a red alert for the next
→24 hours, indicating that the situation might worsen. Schools and junior
→colleges in Mumbai, Thane and Konkan will remain shut on Thursday as a
→precautionary measure."
tags=nltk.pos_tag(nltk.word_tokenize(doc))
list_of_words=['pounding','paralysed','were','turned','deluged','delaying','jamming','flooding']
print([wl.lemmatize(word, 'v') for word in list_of_words])

```

```

['pound', 'paralyse', 'be', 'turn', 'deluge', 'delay', 'jam', 'flood', 'have',
'grind', 'cancel', 'push', 'look', 'reach', 'abandon', 'indicate', 'declare',
'worsen', 'remain']

```

[]: