AIM : Write a python program for natural language processing

Program

import nltk

nltk.download()

new="The big cat ate the little mouse who was after the fresh cheese"

new\_tokens=nltk.word\_tokenize(new)

print(new\_tokens)

new\_tag=nltk.pos\_tag(new\_tokens)

print(new\_tag)

grammer=r"NP: {<DT>?<JJ>\*<NN>}"

chunkParser=nltk.RegexpParser(grammer)

chunked=chunkParser.parse(new\_tag)

print(chunked)

chunked.draw()

Output

['The', 'big', 'cat', 'ate', 'the', 'little', 'mouse', 'who', 'was', 'after', 'the', 'fresh', 'cheese']

[('The', 'DT'), ('big', 'JJ'), ('cat', 'NN'), ('ate', 'VBD'), ('the', 'DT'), ('little', 'JJ'), ('mouse', 'NN'), ('who', 'WP'), ('was', 'VBD'), ('after', 'IN'), ('the', 'DT'), ('fresh', 'JJ'), ('cheese', 'NN')]

(S

(NP The/DT big/JJ cat/NN)

ate/VBD

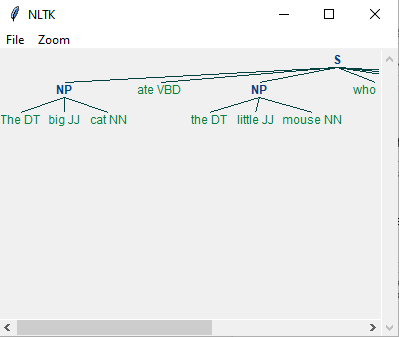
(NP the/DT little/JJ mouse/NN)

who/WP

was/VBD

after/IN

(NP the/DT fresh/JJ cheese/NN))



Program

import nltk

nltk.download('averaged\_perceptron\_tagger')

sample\_text= """

Rama killed ravana to save sita from lanka.the legend of the ramayanam is the most popular

indian epic.a lot of movies and

serials have already been shot in several languages.here in india based on the ramayana

"""

tokenized=nltk.sent\_tokenize(sample\_text)

for i in tokenized:

words=nltk.word\_tokenize(i)

tagged\_words=nltk.pos\_tag(words)

chunkGrams=r"""VB: {}"""

chunkParser=nltk.RegexpParser(chunkGrams)

chunked=chunkParser.parse(tagged\_words)

print(chunked)

chunked.draw()

Output

(S

Rama/NNP

killed/VBD

ravana/NN

to/TO

save/VB

sita/NN

from/IN

lanka.the/JJ

legend/NN

of/IN

the/DT

ramayanam/NN

is/VBZ

the/DT

most/RBS

popular/JJ

indian/JJ

epic.a/NN

lot/NN

of/IN

movies/NNS

and/CC

serials/NNS

have/VBP

already/RB

been/VBN

shot/VBN

in/IN

several/JJ

languages.here/NNS

in/IN

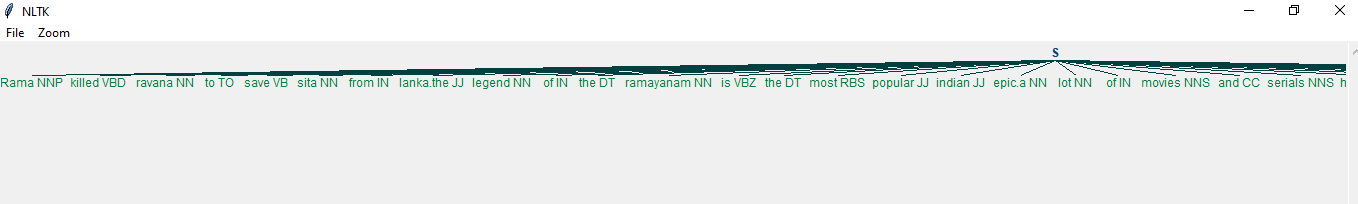
india/NN

based/VBN

on/IN

the/DT

ramayana/NN)



Program

import scrapy

class Quotesspider(scrapy.Spider):

name='quotes'

def start\_requests(self):

urls=['http://quotes.toscrap.com/page/1/'

'http://quotes.toscrap.com/page/2/']

def parse(self, response):

page=response.url.split("/") [-2]

filename='quotes-%s.html' % page

with open(filename, 'web') as f:

f.write(response.body)

self.log('saved file%s' % filename)