Excercise:1 Simple\_Frequency\_Distribution

Code:

import nltk

from nltk.tokenize import TreebankWordTokenizer

from nltk.probability import FreqDist

import matplotlib.pyplot as plt

tokenizer=TreebankWordTokenizer()

text='Natural Language Processing(NLP) with python is fun and educational.python simplifies NLP tasks.'

tokens=tokenizer.tokenize(text)

print("Tokens:",tokens)

fdist=FreqDist(tokens)

print("\n Word Frequency Distribution:")

for word,freq in fdist.items():

print(f"{word}: {freq}")

fdist.plot(title='Simple Frequency Distribution plot')

plt.show()

plt.hist(tokens,bins=10,color="green",edgecolor="blue")

plt.title("Histogram")

plt.show()

Output:

Word Frequency Distribution:

Natural: 1

Language: 1

Processing: 1

(: 1

NLP: 2

): 1

with: 1

python: 1

is: 1

fun: 1

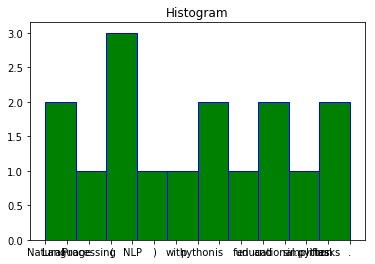
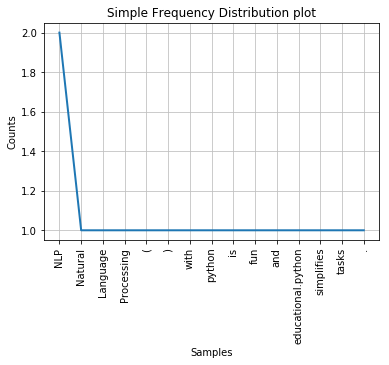
and: 1

educational.python: 1

simplifies: 1

tasks: 1

.: 1



Excercise:2 \_Advanced\_Frequency\_Distribution\_with\_STOPWORDS\_Removal

Code:

import nltk

from nltk.tokenize import TreebankWordTokenizer

from nltk.corpus import stopwords

from nltk.probability import FreqDist

import matplotlib.pyplot as plt

nltk.download('stopwords')

tokenizer=TreebankWordTokenizer()

stop\_words=set(stopwords.words('english'))

text='Natural Language Processing with python is intresting and useful.python makes NLP tasks easier and more efficient'

tokens=tokenizer.tokenize(text)

filtered\_tokens=[Words for Words in tokens if Words.lower() not in stop\_words and Words.isalpha()]

print("filtered\_tokens(withou t stopwords):",filtered\_tokens)

fdist=FreqDist(filtered\_tokens)

print("\n Word Frequency Distribution(without stopwords):")

for word,freq in fdist.items():

print(f"{word}: {freq}")

fdist.plot(title="Word Frequency Distribution(without stopwords)")

plt.hist(tokens,bins=10,color='red',edgecolor="black")

plt.title("Histogram")

plt.show()

output:

filtered\_tokens(withou t stopwords): ['Natural', 'Language', 'Processing', 'python', 'intresting', 'makes', 'NLP', 'tasks', 'easier', 'efficient']

Word Frequency Distribution(without stopwords):

Natural: 1

Language: 1

Processing: 1

python: 1

intresting: 1

makes: 1

NLP: 1

tasks: 1

easier: 1

efficient: 1

