

```
In [1]: import nltk
import string, re
from collections import Counter
from nltk.tokenize import sent_tokenize
from nltk.tokenize import word_tokenize
from nltk.probability import FreqDist
import matplotlib.pyplot as plt
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.tokenize import sent_tokenize, word_tokenize
from nltk.stem.wordnet import WordNetLemmatizer
from nltk.sentiment.vader import SentimentIntensityAnalyzer
import pandas as pd
from textblob import TextBlob
import seaborn as sns
%matplotlib inline
```

```
In [2]: article text that is being used as data source.
"""Imagine a society in which you are rated by the government on your trustworthiness. Your "citizen score" follows you wherever
When you step outside your door, your actions in the physical world are also swept into the dragnet: The government gathers an e
This society may seem dystopian, but it isn't farfetched: It may be China in a few years. The country is racing to become the fi
China's evolving algorithmic surveillance system will rely on the security organs of the communist party-state to filter, collec
ing the text paragraph into sentences
tokenized_text = sent_tokenize(text)
# the broken down paragraph that has been tokenized with sentences
tokenized_text
```

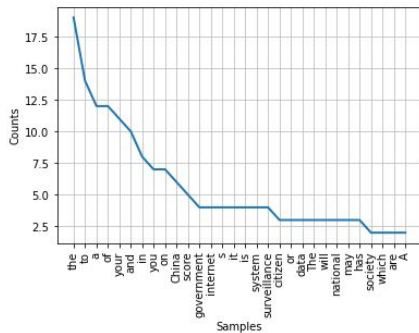
```
['Imagine a society in which you are rated by the government on your trustworthiness.', 'Your "citizen score" follows you where
ver you go.', 'A high score allows you access to faster internet service or a fast-tracked visa to Europe.', 'If you make polit
ical posts online without a permit, or question or contradict the government's official narrative on current events, however, y
our score decreases.', 'To calculate the score, private companies working with your government constantly trawl through vast am
ounts of your social media and online shopping data.', 'When you step outside your door, your actions in the physical world are
also swept into the dragnet: The government gathers an enormous collection of information through the video cameras placed on y
our street and all over your city.', 'If you commit a crime—or simply jaywalk—facial recognition algorithms will match video fo
otage of your face to your photo in a national ID database.', 'It won't be long before the police show up at your door.', 'This
society may seem dystopian, but it isn't farfetched: It may be China in a few years.', 'The country is racing to become the fir
st to implement a pervasive system of algorithmic surveillance.', 'Harnessing advances in artificial intelligence and data mini
ng and storage to construct detailed profiles on all citizens, China's communist party-state is developing a "citizen score" to
incentivize "good" behavior.', 'A vast accompanying network of surveillance cameras will constantly monitor citizens' movemen
ts, purportedly to reduce crime and terrorism.', 'While the expanding Orwellian eye may improve "public safety," it poses a chil
ling new threat to civil liberties in a country that already has one of the most oppressive and controlling governments in the
world.', 'China's evolving algorithmic surveillance system will rely on the security organs of the communist party-state to fil
ter, collect, and analyze staggering volumes of data flowing across the internet.', 'Justifying controls in the name of nationa
l security and social stability, China originally planned to develop what it called a "Golden Shield" surveillance system allow
ing easy access to local, national, and regional records on each citizen.', 'This ambitious project has so far been mostly conf
ined to a content-filtering Great Firewall, which prohibits foreign internet sites including Google, Facebook, and The New York
Times.', 'According to Freedom House, China's level of internet freedom is already the worst on the planet.', 'Now, the Communi
st Party of China is finally building the extensive, multilevel data-gathering system it has dreamed of for decades.']
```

```
In [3]: # Break down the paragraph into single words
tokenized_word = word_tokenize(text)
# Remove all punctuations
clean_tokenized_words = [tokenized_word for tokenized_word in tokenized_word if tokenized_word.isalnum()]
# Print out the results of breaking down the paragraph into single words
print(clean_tokenized_words)
```

```
[ 'Imagine', 'a', 'society', 'in', 'which', 'you', 'are', 'rated', 'by', 'the', 'government', 'on', 'your', 'trustworthiness',
'Your', 'citizen', 'score', 'follows', 'you', 'wherever', 'you', 'go', 'A', 'high', 'score', 'allows', 'you', 'access', 'to',
'faster', 'internet', 'service', 'or', 'a', 'visa', 'to', 'Europe', 'If', 'you', 'make', 'political', 'posts', 'online', 'witho
ut', 'a', 'permit', 'or', 'question', 'or', 'contradict', 'the', 'government', 's', 'official', 'narrative', 'on', 'current',
'events', 'however', 'your', 'score', 'decreases', 'To', 'calculate', 'the', 'score', 'private', 'companies', 'working', 'wit
h', 'your', 'government', 'constantly', 'trawl', 'through', 'vast', 'amounts', 'of', 'your', 'social', 'media', 'and', 'onlin
e', 'shopping', 'data', 'when', 'you', 'step', 'outside', 'your', 'door', 'your', 'actions', 'in', 'the', 'physical', 'world',
'are', 'also', 'swept', 'into', 'the', 'dragnet', 'The', 'government', 'gathers', 'an', 'enormous', 'collection', 'of', 'inform
ation', 'through', 'the', 'video', 'cameras', 'placed', 'on', 'your', 'street', 'and', 'all', 'over', 'your', 'city', 'If', 'yo
u', 'commit', 'a', 'simply', 'recognition', 'algorithms', 'will', 'match', 'video', 'footage', 'of', 'your', 'face', 'to', 'you
r', 'photo', 'in', 'a', 'national', 'ID', 'database', 'It', 'won', 't', 'be', 'long', 'before', 'the', 'police', 'show', 'up',
'at', 'your', 'door', 'This', 'society', 'may', 'seem', 'dystopian', 'but', 'it', 'isn', 't', 'farfetched', 'It', 'may', 'be',
'China', 'in', 'a', 'few', 'years', 'The', 'country', 'is', 'racing', 'to', 'become', 'the', 'first', 'to', 'implement', 'a',
'pervasive', 'system', 'of', 'algorithmic', 'surveillance', 'Harnessing', 'advances', 'in', 'artificial', 'intelligence', 'an
d', 'data', 'mining', 'and', 'storage', 'to', 'construct', 'detailed', 'profiles', 'on', 'all', 'citizens', 'China', 's', 'comm
unist', 'is', 'developing', 'a', 'citizen', 'score', 'to', 'incentivize', 'good', 'behavior', 'A', 'vast', 'accompanying', 'net
work', 'of', 'surveillance', 'cameras', 'will', 'constantly', 'monitor', 'citizens', 'movements', 'purportedly', 'to', 'reduc
e', 'crime', 'and', 'terrorism', 'while', 'the', 'expanding', 'Orwellian', 'eye', 'may', 'improve', 'public', 'safety', 'it',
'poses', 'a', 'chilling', 'new', 'threat', 'to', 'civil', 'liberties', 'in', 'a', 'country', 'that', 'already', 'has', 'one',
'of', 'the', 'most', 'oppressive', 'and', 'controlling', 'governments', 'in', 'the', 'world', 'China', 's', 'evolving', 'algori
thmic', 'surveillance', 'system', 'will', 'rely', 'on', 'the', 'security', 'organs', 'of', 'the', 'communist', 'to', 'filter',
'collect', 'and', 'analyze', 'staggering', 'volumes', 'of', 'data', 'flowing', 'across', 'the', 'internet', 'Justifying', 'cont
rols', 'in', 'the', 'name', 'of', 'national', 'security', 'and', 'social', 'stability', 'China', 'originally', 'planned', 'to',
'develop', 'what', 'it', 'called', 'a', 'Golden', 'Shield', 'surveillance', 'system', 'allowing', 'easy', 'access', 'to', 'loca
l', 'national', 'and', 'regional', 'records', 'on', 'each', 'citizen', 'This', 'ambitious', 'project', 'has', 'so', 'far', 'bee
n', 'mostly', 'confined', 'to', 'a', 'Great', 'Firewall', 'which', 'prohibits', 'foreign', 'internet', 'sites', 'including', 'G
oogle', 'Facebook', 'and', 'The', 'New', 'York', 'Times', 'According', 'to', 'Freedom', 'House', 'China', 's', 'level', 'of',
'internet', 'freedom', 'is', 'already', 'the', 'worst', 'on', 'the', 'planet', 'Now', 'the', 'Communist', 'Party', 'of', 'Chin
a', 'is', 'finally', 'building', 'the', 'extensive', 'multilevel', 'system', 'it', 'has', 'dreamed', 'of', 'for', 'decades']
```

```
In [4]: # Declare a frequency distribution with the final version of the tokenized paragraphs
frequency_distribution = FreqDist(clean_tokenized_words)

# Plot the distribution of words used in the article
frequency_distribution.plot(30, cumulative = False)
# Display the graph on frequency distribution
plt.show()
```



```
In [5]: # Find all the stop words/noise words/words not needed for sentiment in the article
stop_words_removal = set(stopwords.words("english"))
# Display all the stop words found in the article
print(stop_words_removal)
```

```
{ 'some', 'up', 'he', 'myself', 'how', 'by', 's', 'nor', 'both', "hasn't", "don't", 'you', 'and', 'wouldn', 'very', 'll', 'am',
'if', 'into', "you'd", 'your', 'more', "that'll", 'y', 'so', 'again', 'there', "you'll", "aren't", 'shan', 'which', 'no', 'i',
'hasn', 'but', 'its', 'out', 'such', 'ourselves', 'too', 'who', 'before', 'my', 'of', 'while', 'can', 'm', 'from', 'haven', "yo
u've", "shouldn't", 'themselves', 'will', "didn't", 'was', 'you're', "doesn't", 've', 'are', 'under', 'over', 'their', 'only',
'hadn', 'where', 'those', 'against', 'that', 'after', 'few', 'not', 'mustn', "won't", 'on', 'did', 'these', 'she', 'what', 'th
e', 't', 'has', 'don', 'other', 'why', 'until', 'does', 'during', 'in', 'because', 'didn', 'hers', 'about', 'needn', 'him', 'mi
ghtn', 'be', 'yourself', 'itself', 'most', 'than', 'ain', 'have', 'couldn', 'been', 'each', 'to', "hadn't", "isn't", 'wasn', 'i
sn', 'whom', 'at', 'through', 'her', 'between', 'had', 'same', 'for', "wouldn't", 'having', 'yours', 'do', 'we', 'as', 'off',
'once', 'shouldn', 'were', 'being', 'herself', "it's", 'here', 'all', "shan't", 'won', 'down', 'an', 'should', 'is', 'our', 'ab
ove', 'when', 'doing', "couldn't", 'doesn', "should've", 'yourselves', 'it', 'himself', "needn't", 'on', 'with', 'ours', 'the
y', 'o', 'she's', "mightn't", 'theirs', "weren't", 'this', 'weren', 'd', 'me', 'a', 'any', "haven't", 'ma', 'his', 'aren', "was
n't", 'own', 'just', 'below', 'them', 'then', 're', "mustn't", 'now', 'further'}
```

```
In [6]: # Filter out all the stop words
# Create a list of what the article would look like with the stop words gone
filtered_stop_words = []
# for loop for every word found in the clean_tokenized_words list
for w in clean_tokenized_words:
    # If statement for the condition of if a word is not in the stop_words_removal list then append the word
    if w not in stop_words_removal:
        # Append: Adding an item to the end of the filtered_stop_words list
        filtered_stop_words.append(w)
# Display the clean_tokenized_words list
print("Tokenized Sentence:",clean_tokenized_words)
# Display the filtered_stop_words list to compare
print("Filtered Sentence:",filtered_stop_words)
```

```
Tokenized Sentence: ['Imagine', 'a', 'society', 'in', 'which', 'you', 'are', 'rated', 'by', 'the', 'government', 'on', 'your',
'trustworthiness', 'Your', 'citizen', 'score', 'follows', 'you', 'wherever', 'you', 'go', 'A', 'high', 'score', 'allows', 'yo
u', 'access', 'to', 'faster', 'internet', 'service', 'on', 'a', 'visa', 'to', 'Europe', 'If', 'you', 'make', 'political', 'post
s', 'online', 'without', 'a', 'permit', 'or', 'question', 'on', 'contradict', 'the', 'government', 's', 'official', 'narrativ
e', 'on', 'current', 'events', 'however', 'your', 'score', 'decreases', 'To', 'calculate', 'the', 'score', 'private', 'companie
s', 'working', 'with', 'your', 'government', 'constantly', 'trawl', 'through', 'vast', 'amounts', 'of', 'your', 'social', 'medi
a', 'and', 'online', 'shopping', 'data', 'when', 'you', 'step', 'outside', 'your', 'door', 'your', 'actions', 'in', 'the', 'phy
sical', 'world', 'are', 'also', 'swept', 'into', 'the', 'dragnet', 'The', 'government', 'gathers', 'an', 'enormous', 'collectio
n', 'of', 'information', 'through', 'the', 'video', 'cameras', 'placed', 'on', 'your', 'street', 'and', 'all', 'over', 'your',
'city', 'If', 'you', 'commit', 'a', 'simply', 'recognition', 'algorithms', 'will', 'match', 'video', 'footage', 'of', 'your',
'face', 'to', 'your', 'photo', 'in', 'a', 'national', 'ID', 'database', 'It', 'won', 't', 'be', 'long', 'before', 'the', 'police',
'e', 'show', 'up', 'at', 'your', 'door', 'This', 'society', 'may', 'seem', 'dystopian', 'but', 'it', 'isn', 't', 'farfetched',
'It', 'may', 'be', 'China', 'in', 'a', 'few', 'years', 'The', 'country', 'is', 'racing', 'to', 'become', 'the', 'first', 'to',
'implement', 'a', 'pervasive', 'system', 'of', 'algorithmic', 'surveillance', 'Harnessing', 'advances', 'in', 'artificial', 'in
telligence', 'and', 'data', 'mining', 'and', 'storage', 'to', 'construct', 'detailed', 'profiles', 'on', 'all', 'citizens', 'Ch
ina', 's', 'communist', 'is', 'developing', 'a', 'citizen', 'score', 'to', 'incentivize', 'good', 'behavior', 'A', 'vast', 'acc
ompanying', 'network', 'of', 'surveillance', 'cameras', 'will', 'constantly', 'monitor', 'citizens', 'movements', 'purportedl
y', 'to', 'reduce', 'crime', 'and', 'terrorism', 'While', 'the', 'expanding', 'Orwellian', 'eye', 'may', 'improve', 'public',
'safety', 'it', 'poses', 'a', 'chilling', 'new', 'threat', 'to', 'civil', 'liberties', 'in', 'a', 'country', 'that', 'already',
'has', 'one', 'of', 'the', 'most', 'oppressive', 'and', 'controlling', 'governments', 'in', 'the', 'world', 'China', 's', 'evol
ving', 'algorithmic', 'surveillance', 'system', 'will', 'rely', 'on', 'the', 'security', 'organs', 'of', 'the', 'communist', 't
o', 'filter', 'collect', 'and', 'analyze', 'staggering', 'volumes', 'of', 'data', 'flowing', 'across', 'the', 'internet', 'Just
ifying', 'controls', 'in', 'the', 'name', 'of', 'national', 'security', 'and', 'social', 'stability', 'China', 'originally', 'p
lanned', 'to', 'develop', 'what', 'it', 'called', 'a', 'Golden', 'Shield', 'surveillance', 'system', 'allowing', 'easy', 'acce
s', 'to', 'local', 'national', 'and', 'regional', 'records', 'on', 'each', 'citizen', 'This', 'ambitious', 'project', 'has', 's
o', 'far', 'been', 'mostly', 'confined', 'to', 'a', 'Great', 'Firewall', 'which', 'prohibits', 'foreign', 'internet', 'sites',
'including', 'Google', 'Facebook', 'and', 'The', 'New', 'York', 'Times', 'According', 'to', 'Freedom', 'House', 'China', 's',
'level', 'of', 'internet', 'freedom', 'is', 'already', 'the', 'worst', 'on', 'the', 'planet', 'Now', 'the', 'Communist', 'Part
y', 'of', 'China', 'is', 'finally', 'building', 'the', 'extensive', 'multilevel', 'system', 'it', 'has', 'dreamed', 'of', 'fo
r', 'decades']
```

```
Filtered Sentence: ['Imagine', 'society', 'rated', 'government', 'trustworthiness', 'Your', 'citizen', 'score', 'follows', 'wher
ever', 'go', 'A', 'high', 'score', 'allows', 'access', 'faster', 'internet', 'service', 'visa', 'Europe', 'If', 'make', 'politi
cal', 'posts', 'online', 'without', 'permit', 'question', 'contradict', 'government', 'official', 'narrative', 'current', 'even
ts', 'however', 'score', 'decreases', 'To', 'calculate', 'score', 'private', 'companies', 'working', 'government', 'constantly',
y', 'trawl', 'vast', 'amounts', 'social', 'media', 'online', 'shopping', 'data', 'when', 'step', 'outside', 'door', 'actions',
'physical', 'world', 'also', 'swept', 'dragnet', 'The', 'government', 'gathers', 'enormous', 'collection', 'information', 'vide
o', 'cameras', 'placed', 'street', 'city', 'If', 'commit', 'simply', 'recognition', 'algorithms', 'match', 'video', 'footage',
'face', 'photo', 'national', 'ID', 'database', 'It', 'long', 'police', 'show', 'door', 'society', 'may', 'seem', 'dysto
pian', 'farfetched', 'It', 'may', 'China', 'years', 'The', 'country', 'racing', 'become', 'first', 'implement', 'pervasive', 's
ystem', 'algorithmic', 'surveillance', 'Harnessing', 'advances', 'artificial', 'intelligence', 'data', 'mining', 'storage', 'co
nstruct', 'detailed', 'profiles', 'citizens', 'China', 'communist', 'developing', 'citizen', 'score', 'incentivize', 'good', 'b
ehavior', 'A', 'vast', 'accompanying', 'network', 'surveillance', 'cameras', 'constantly', 'monitor', 'citizens', 'movements',
'purportedly', 'reduce', 'crime', 'terrorism', 'While', 'expanding', 'Orwellian', 'eye', 'may', 'improve', 'public', 'safety',
'poses', 'chilling', 'new', 'threat', 'civil', 'liberties', 'country', 'already', 'one', 'oppressive', 'controlling', 'governme
nts', 'world', 'China', 'evolving', 'algorithmic', 'surveillance', 'system', 'rely', 'security', 'organs', 'communist', 'filte
r', 'collect', 'analyze', 'staggering', 'volumes', 'data', 'flowing', 'across', 'internet', 'Justifying', 'controls', 'name',
'national', 'security', 'social', 'stability', 'China', 'originally', 'planned', 'develop', 'called', 'Golden', 'Shield', 'surv
eillance', 'system', 'allowing', 'easy', 'access', 'local', 'national', 'regional', 'records', 'citizen', 'This', 'ambitious',
'project', 'far', 'mostly', 'confined', 'Great', 'Firewall', 'prohibits', 'foreign', 'internet', 'sites', 'including', 'Googl
e', 'Facebook', 'The', 'New', 'York', 'Times', 'According', 'Freedom', 'House', 'China', 'level', 'internet', 'freedom', 'alrea
dy', 'worst', 'planet', 'Now', 'Communist', 'Party', 'China', 'finally', 'building', 'extensive', 'multilevel', 'system', 'drea
med', 'decades']
```

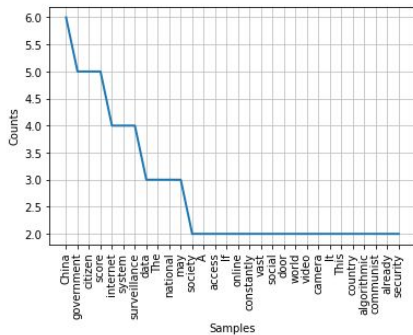
```
In [7]: # Lemmatization: Reduce the words to their base word, removing any synonyms
# Create a list of what the article would look like after removing stop words and being lemmatized
lemmatized_words = []
# for loop for every word found in the filtered_stop_words list
for w in filtered_stop_words:
    # Lemmatize the word
    w = WordNetLemmatizer().lemmatize(w)
    # Append: Adding an item to the end of the lemmatized_words list
    lemmatized_words.append(w)
# Display the filtered_stop_words
print("Filtered Sentence:",filtered_stop_words)
# Display the Lemmatized_words list to compare
print("Lemmatized Word:",lemmatized_words)
```

Filtered Sentence: ['Imagine', 'society', 'rated', 'government', 'trustworthiness', 'Your', 'citizen', 'score', 'follows', 'where', 'rever', 'go', 'A', 'high', 'score', 'allows', 'access', 'faster', 'internet', 'service', 'visa', 'Europe', 'If', 'make', 'political', 'posts', 'online', 'without', 'permit', 'question', 'contradict', 'government', 'official', 'narrative', 'current', 'events', 'however', 'score', 'decreases', 'To', 'calculate', 'score', 'private', 'companies', 'working', 'government', 'constantly', 'y', 'trawl', 'vast', 'amounts', 'social', 'media', 'online', 'shopping', 'data', 'When', 'step', 'outside', 'door', 'actions', 'physical', 'world', 'also', 'swept', 'dragnet', 'The', 'government', 'gathers', 'enormous', 'collection', 'information', 'video', 'cameras', 'placed', 'street', 'city', 'If', 'commit', 'simply', 'recognition', 'algorithms', 'match', 'video', 'footage', 'face', 'photo', 'national', 'ID', 'database', 'It', 'long', 'police', 'show', 'door', 'This', 'society', 'may', 'seem', 'dystopian', 'farfetched', 'It', 'may', 'China', 'year', 'The', 'country', 'racing', 'become', 'first', 'implement', 'pervasive', 'system', 'algorithmic', 'surveillance', 'Harnessing', 'advances', 'artificial', 'intelligence', 'data', 'mining', 'storage', 'construct', 'detailed', 'profiles', 'citizens', 'China', 'communist', 'developing', 'citizen', 'score', 'incentivize', 'good', 'behavior', 'A', 'vast', 'accompanying', 'network', 'surveillance', 'cameras', 'constantly', 'monitor', 'citizens', 'movements', 'purportedly', 'reduce', 'crime', 'terrorism', 'While', 'expanding', 'Orwellian', 'eye', 'may', 'improve', 'public', 'safety', 'poses', 'chilling', 'new', 'threat', 'civil', 'liberties', 'country', 'already', 'one', 'oppressive', 'controlling', 'governments', 'world', 'China', 'evolving', 'algorithmic', 'surveillance', 'system', 'rely', 'security', 'organ', 'communist', 'filter', 'collect', 'analyze', 'staggering', 'volumes', 'data', 'flowing', 'across', 'internet', 'Justifying', 'controls', 'name', 'national', 'security', 'social', 'stability', 'China', 'originally', 'planned', 'develop', 'called', 'Golden', 'Shield', 'surveillance', 'system', 'allowing', 'easy', 'access', 'local', 'national', 'regional', 'records', 'citizen', 'This', 'ambitious', 'project', 'far', 'mostly', 'confined', 'Great', 'Firewall', 'prohibits', 'foreign', 'internet', 'sites', 'including', 'Google', 'Facebook', 'The', 'New', 'York', 'Times', 'According', 'Freedom', 'House', 'China', 'level', 'internet', 'freedom', 'already', 'worst', 'planet', 'Now', 'Communist', 'Party', 'China', 'finally', 'building', 'extensive', 'multilevel', 'system', 'dreamed', 'decades']

Lemmaized Word: ['Imagine', 'society', 'rated', 'government', 'trustworthiness', 'Your', 'citizen', 'score', 'follows', 'where', 'ver', 'go', 'A', 'high', 'score', 'allows', 'access', 'faster', 'internet', 'service', 'visa', 'Europe', 'If', 'make', 'political', 'post', 'online', 'without', 'permit', 'question', 'contradict', 'government', 'official', 'narrative', 'current', 'event', 'however', 'score', 'decrease', 'To', 'calculate', 'score', 'private', 'company', 'working', 'government', 'constantly', 'trawl', 'vast', 'amount', 'social', 'medium', 'online', 'shopping', 'data', 'When', 'step', 'outside', 'door', 'action', 'physical', 'world', 'also', 'swept', 'dragnet', 'The', 'government', 'gather', 'enormous', 'collection', 'information', 'video', 'camera', 'placed', 'street', 'city', 'If', 'commit', 'simply', 'recognition', 'algorithm', 'match', 'video', 'footage', 'face', 'photo', 'national', 'ID', 'database', 'It', 'long', 'police', 'show', 'door', 'This', 'society', 'may', 'seem', 'dystopian', 'farfetched', 'It', 'may', 'China', 'year', 'The', 'country', 'racing', 'become', 'first', 'implement', 'pervasive', 'system', 'algorithmic', 'surveillance', 'Harnessing', 'advance', 'artificial', 'intelligence', 'data', 'mining', 'storage', 'construct', 'detailed', 'profile', 'citizen', 'China', 'communist', 'developing', 'citizen', 'score', 'incentivize', 'good', 'behavior', 'A', 'vast', 'accompanying', 'network', 'surveillance', 'camera', 'constantly', 'monitor', 'citizen', 'movement', 'purportedly', 'reduce', 'crime', 'terrorism', 'While', 'expanding', 'Orwellian', 'eye', 'may', 'improve', 'public', 'safety', 'pose', 'chilling', 'new', 'threat', 'civil', 'liberty', 'country', 'already', 'one', 'oppressive', 'controlling', 'government', 'world', 'China', 'evolving', 'algorithmic', 'surveillance', 'system', 'rely', 'security', 'organ', 'communist', 'filter', 'collect', 'analyze', 'staggering', 'volume', 'data', 'flowing', 'across', 'internet', 'Justifying', 'control', 'name', 'national', 'security', 'social', 'stability', 'China', 'originally', 'planned', 'develop', 'called', 'Golden', 'Shield', 'surveillance', 'system', 'allowing', 'easy', 'access', 'local', 'national', 'regional', 'record', 'citizen', 'This', 'ambitious', 'project', 'far', 'mostly', 'confined', 'Great', 'Firewall', 'prohibits', 'foreign', 'internet', 'site', 'including', 'Google', 'Facebook', 'The', 'New', 'York', 'Times', 'According', 'Freedom', 'House', 'China', 'level', 'internet', 'freedom', 'already', 'worst', 'planet', 'Now', 'Communist', 'Party', 'China', 'finally', 'building', 'extensive', 'multilevel', 'system', 'dreamed', 'decade']

```
In [8]: # Declare a new frequency distribution for the final version of the article list
frequency_distribution_cleaned = FreqDist(lemmatized_words)

# Plot the distribution of words after it's been through tokenization, filtered stop words, and Lemmaized
frequency_distribution_cleaned.plot(30, cumulative = False)
# Display the final version of the graph on frequency distribution
plt.show()
```



```
In [9]: # Function to convert the Lemmatized_words List into strings
def listToString(s):

    # Initializing an empty string
    string1 = ""

    # Return the string values
    return (string1.join(s))
# Put all the words in the Lemmatized_words List into the ListToString function
s = lemmatized_words
# Print out the results
print(listToString(s))
```

Imagine society rated government trustworthiness Your citizen score follows wherever go A high score allows access faster internet service visa Europe If make political post online without permit question contradict government official narrative current event however score decrease To calculate score private company working government constantly trawl vast amount social medium online shopping data When step outside door action physical world also swept dragnet The government gather enormous collection information video camera placed street city If commit simply recognition algorithm match video footage face photo national ID database It long police show door This society may seem dystopian farfetched It may China year The country racing become first implement pervasive system algorithmic surveillance Harnessing advance artificial intelligence data mining storage construct detailed profile citizen China communist developing citizen score incentivize good behavior A vast accompanying network surveillance camera constantly monitor citizen movement purportedly reduce crime terrorism While expanding Orwellian eye may improve public safety pose chilling new threat civil liberty country already one oppressive controlling government world China evolving algorithmic surveillance system rely security organ communist filter collect analyze staggering volume data flowing across internet Justifying control name national security social stability China originally planned develop called Golden Shield surveillance system allowing easy access local national regional record citizen This ambitious project far mostly confined Great Firewall prohibits foreign internet site including Google Facebook The New York Times According Freedom House China level internet freedom already worst planet Now Communist Party China finally building extensive multilevel system dreamed decade

```
In [22]: # Using the TextBlob library for text sentiment analysis which will display the sentiment polarity and subjectivity
# Polarity range: [-1.0,1.0] | -1 = negative sentiment | 1 = positive sentiment
# Have the TextBlob library analyze the scrubbed article
text_analyze = TextBlob("Imagine society rated government trustworthiness Your citizen score follows wherever go A high score all
# Calculate the sentiment polarity
text_analyze.sentiment.polarity

# Display the sentiment polarity
print("This is the sentiment polarity: ", text_analyze.sentiment.polarity)

# Variable holding data on the calculated sentiment polarity
a = text_analyze.sentiment.polarity

# Condition for if the article has a positive sentiment
if a > 0.50:
    print("This Article has a Positive Sentiment")
# Condition for if the article has a negative sentiment
elif a < 0.50:
    print("This article has a Negative Sentiment")
# Condition for if the article has a neutral sentiment
else:
    print("This article has a Neutral Sentiment")
```

```
This is the sentiment polarity: 0.06757575757575758
This article has a Negative Sentiment
```