

In [26]:



```
import docx2txt
import seaborn as sns
```

In [5]:



```
path = 'testdoc.docx'
text = docx2txt.process(path)
```

In [9]:



```
print(text)
```

Millions of people in India took part in an annual tree planting drive Sunday. More than 250 million saplings were planted in a single day across the country's most-populous state.

The campaign was led by Uttar Pradesh state government officials, lawmakers, and activists, in a bid to reduce carbon emissions and combat climate change.

Where were the trees planted?

The saplings were planted by volunteers in forests, farms, schools, and along riverbanks and highways.

"We are committed to increasing the forest cover of Uttar Pradesh to over 15% of the total land area in the next five years," said state forest official Manoj Singh.

According to another government official, the forest cover of the state has increased over the last few years.

"There has been an increase of 127 square kilometers [79 square miles] in the forest cover in Uttar Pradesh as compared to 2017," a state government spokesperson was quoted as saying in The Indian Express newspaper.

"There has also been an increase in trees and plants. The tree cover has increased to 3.05%, as compared to the national average of 2.89%," the official said, citing the 2019 Forest Survey of India report.

How many saplings survive?

Uttar Pradesh State Forest Minister Dara Singh said the long-term survival of the saplings remains a concern, adding that usually only 60% of the saplings survive. The rest succumb to disease or lack of water.

However, he said that about 80% of the saplings planted in the last four annual drives have survived.

"All the regions where plantation is being carried out have been geo-tagged so that we can ascertain what exactly happened," Chauhan told The Pioneer newspaper.

"These saplings carry QR codes so that officials can maintain a record and verify whether the saplings survived or not. Besides, teams have been formed to monitor progress of the plantation drive," he said.

What is the extent of India's tree planting project?

India has vowed to have a third of its total land area, or 95 million hectares, under forest and tree cover by 2030.

The government has allocated \$6.2 billion (€5.2 billion) for the tree-planting across the country.

However, industrial development and a rapidly growing population has put further stress on the land.

In [10]:



```
path2 = 'testdoc2.docx'  
text2 = docx2txt.process(path2)  
print(text2)
```


Between 2016 and 2019, the state forest department under the BJP government had launched 'Green Maharashtra' drive with an aim to plant 50 crore trees across the state in the four-year period. In October 2019, the government had claimed it had surpassed the target by planting 33 crore trees in July-September 2019. The Indian Express had found that non-forest agencies – such as gram panchayats – which were tasked with planting trees had not uploaded the mandatory audio-visual proof of the tree plantation drives on the specially created portal.

In Pune Revenue Division, it was claimed the gram panchayats planted 1.7 crore saplings; however, no evidence was uploaded for 87 per cent (1.49 crore) saplings. Also, out of the 59 government agencies involved in the drive as many as 38 had not submitted survival reports about the saplings.

This year, the targets set by the forest department were comparatively modest. For example, Pune Circle – which comprises three divisions in Pune and Solapur district – had planned to plant 17 lakh saplings on forest land. However, it may not meet the target due to unavailability of funds. Last year Pune Circle planted 70 lakh saplings on forest land.

In Pune Division – which comprises six talukas namely Maval, Mulshi, Daund, Indapur, Baramati and Havveli – preparations were done for plantation of about 4 lakh trees with special emphasis on teakwood.

The National Forest Policy aims and emphasizes at maintaining 33% of the country's geographical area under forest and green cover. In view of this and as a part of the 50 crore plantation programme within Maharashtra, the Maharashtra Forest Department aims to plant 4 crore saplings all over the State between 1st to July 7th, 2017 to celebrate 'Vanmohotsav'.

The plantation programme, which was announced in 2016 with the aim of planting 2 crore trees on 1st July 2016 was a resounding success with the final total reported figure of 2.82 crore saplings planted on a single day. To maintain consistency of this platform without affecting its momentum, the Forest Department has set the target of plantation of 4 crore, 13 crore and 33 crore saplings under the mission of 50 crore plantation which shall be accomplished in the three consecutive years viz. 2017, 2018 and 2019.

The 4 crore saplings for the year 2017 will be planted during the Vanmohotsav, July 1st to July 7th in a state-wide drive with the involvement of 33 government departments along with Students of Schools and Colleges, NSS, NCC, CSR, NGOs, Railways, National Highways, Defence, NABARD and other stakeholders of Society.

In a first of its kind, a 24-hour toll free helpline number 1926 called 'Hello Forest' has been set up to provide information regarding plantation, protection and for mass awareness. The Forest Department has created a mobile application called 'My Plants' to record details of the plantation such as numbers, species and location into the computer system of the Forest Department. All volunteers at individual, collective and organizational level should download and use this application to record their tree plantation work through the application, which will be operational from 1st July to 7th July.

In [14]:



```
In consonance of the public participation, the Maharashtra Forest Department has initiated the 'Maharashtra Harit Sena'/'Green Army' which is a body of dedicated volunteers to participate in the plantation, protection, and activities in forest, wildlife, and related sectors around the year. Individuals and organisations interested in volunteering can register on the Green Army website www.greenarmy.mahaforest.gov.in
```

An integrated drive has been set in place to ensure seamless and successful participation from all stakeholders of society, especially the public.

['Millions of people in India took part in an annual tree planting drive Sunday.', 'More than 250 million saplings were planted in a single day across the country's most-populous state.', 'The campaign was led by Uttar Pradesh state government officials, lawmakers, and activists, in a bid to reduce carbon emissions and combat climate change.', 'Where were the trees planted?', 'The saplings were planted by volunteers in forests, farms, schools, and along riverbanks and highways.', '"We are committed to increasing the forest cover of Uttar Pradesh to over 15% of the total land area in the next five years,\'' said state forest official Manoj Singh.', 'According to another government official, the forest cover of the state has increased over the last few years.', '"There has been an increase of 127 square kilometers [79 square miles] in the forest cover in Uttar Pradesh as compared to 2017," a state government spokesperson was quoted as saying in The Indian Express newspaper.', '"There has also been an increase in trees and plants.', 'The tree cover has increased to 3.05%, as compared to the national average of 2.89%," the official said, citing the 2019 Forest Survey of India report.', 'How many saplings survive?', 'Uttar Pradesh State Forest Minister Dara Singh said the long-term survival of the saplings remains a concern, adding that usually only 60% of the saplings survive.', 'The rest succumb to disease or lack of water.', 'However, he said that about 80% of the saplings planted in the last four annual drives have survived.', '"All the regions where plantation is being carried out have been geo-tagged so that we can ascertain what exactly happened," Chauhan told The Pioneer newspaper.', '"These saplings carry QR codes so that officials can maintain a record and verify whether the saplings survived or not.', 'Besides, teams have been formed to monitor progress of the plantation drive," he said.', 'What is the extent of India's tree planting project?', 'India has vowed to have a third of its total land area, or 95 million hectares, under forest and tree cover by 2030.', 'The government has allocated \$6.2 billion (€5.2 billion) for the tree-planting across the country.', 'However, industrial development and a rapidly growing population has put further stress on the land.']['Between 2016 and 2019, the state forest department under the BJP government had launched 'Green Maharashtra' drive with an aim to plant 50 crore trees across the state in the four-year period.', 'In October 2019, the government had claimed it had surpassed the target by planting 33 crore trees in July-September 2019.', 'The Indian Express had found that non-forest agencies – such as gram panchayats – which were tasked with planting trees had not uploaded the mandatory audio-visual proof of the tree plantation drives on the specially created portal.', 'In Pune Revenue Division, it was claimed the gram panchayats planted 1.7 crore saplings; however, no evidence was uploaded for 87 per cent (1.49 crore) saplings.', 'Also, out of the 59 government agencies involved in the drive as many as 38 had not submitted survival reports about the saplings.', 'This year, the targets set by the forest department were comparatively modest.', 'For example, Pune Circle – which comprises three divisions in Pune and Solapur district – had planned to plant 17 lakh saplings on forest land.', 'However, it may not meet the target due to unavailability of funds.', 'Last year Pune Circle planted 70 lakh saplings on forest land.', 'In Pune Division – which comprises six talukas namely Maval, Mulshi, Daund, Indapur, Baramati and Havveli – preparations were done for plantation of about 4 lakh trees with special emphasis on teakwood.', 'The National Forest Policy aims and emphasizes at maintaining 33% of the country's geographical area under forest and green cover.', 'In view of this and as a part of the 50 crore plantation programme within Maharashtra, the Maharashtra Forest Department aims to plant 4 crore saplings all over the State between 1st to July 7th, 2017 to celebrate 'Vanmohotsav'.', 'The plantation programme, which was announced in 2016 with the aim of planting 2 crore trees on 1st July 2016 was a resounding success with the final total reported figure of 2.82 crore saplings planted on a single day.', 'To maintain consistency of this platform without affecting its momentum, the Forest Department

has set the target of plantation of 4 crore, 13 crore and 33 crore saplings under the mission of 50 crore plantation which shall be accomplished in the three consecutive years viz.', '2017, 2018 and 2019.', 'The 4 crore saplings for the year 2017 will be planted during the Vanmohotsav, July 1st to July 7th in a state-wide drive with the involvement of 33 government departments along with Students of Schools and Colleges, NSS, NCC, CSR, NGOs, Railways, National Highways, Defence, NABARD and other stakeholders of Society.', 'In a first of its kind, a 24-hour toll free helpline number 1926 called 'Hello Forest' has been set up to provide information regarding plantation, protection and for mass awareness.', 'The Forest Department has created a mobile application called 'My Plants' to record details of the plantation such as numbers, species and location into the computer system of the Forest Department.', 'All volunteers at individual, collective and organizational level should download and use this application to record their tree plantation work through the application, which will be operational from 1st July to 7th July.', 'In consonance of the public participation, the Maharashtra Forest Department has initiated the 'Maharashtra Harit Sena' / 'Green Army' which is a body of dedicated volunteers to participate in the plantation, protection, and activities in forest, wildlife, and related sectors around the year.', 'Individuals and organisations interested in volunteering can register on the Green Army website www.greenarmy.mahaforest.gov.in. An integrated drive has been set in place to ensure seamless and successful participation from all stakeholders of society, especially the public.']



['Millions', 'of', 'people', 'in', 'India', 'took', 'part', 'in', 'an', 'annual', 'tree', 'planting', 'drive', 'Sunday', '.', 'More', 'than', '25 0', 'million', 'saplings', 'were', 'planted', 'in', 'a', 'single', 'day', 'across', 'the', 'country', "'s", 'most-populous', 'state', '.', 'The', 'campaign', 'was', 'led', 'by', 'Uttar', 'Pradesh', 'state', 'government', 'officials', ',', 'lawmakers', ',', 'and', 'activists', ',', 'in', 'a', 'bid', 'to', 'reduce', 'carbon', 'emissions', 'and', 'combat', 'climate', 'change', '.', 'Where', 'were', 'the', 'trees', 'planted', '?', 'The', 'saplings', 'were', 'planted', 'by', 'volunteers', 'in', 'forests', ',', 'farms', ',', 'schools', ',', 'and', 'along', 'riverbanks', 'and', 'highways', '.', 'We', 'are', 'committed', 'to', 'increasing', 'the', 'forest', 'cover', 'of', 'Uttar', 'Pradesh', 'to', 'over', '15', '%', 'of', 'the', 'total', 'land', 'area', 'in', 'the', 'next', 'five', 'years', ',', 'said', 'state', 'forest', 'official', 'Manoj', 'Singh', '.', 'According', 'to', 'another', 'government', 'official', ',', 'the', 'forest', 'cover', 'of', 'the', 'state', 'has', 'increased', 'over', 'the', 'last', 'few', 'years', '.', 'There', 'has', 'been', 'an', 'increase', 'of', '127', 'square', 'kilometers', '[', '79', 'square', 'miles', ']', 'in', 'the', 'forest', 'cover', 'in', 'Uttar', 'Pradesh', 'as', 'compared', 'to', '2017', ',', 'a', 'state', 'government', 'spokesperson', 'was', 'quoted', 'as', 'saying', 'in', 'The', 'Indian', 'Express', 'newspaper', '.', 'There', 'has', 'also', 'been', 'an', 'increase', 'in', 'trees', 'and', 'plants', '.', 'The', 'tree', 'cover', 'has', 'increased', 'to', '3.05', '%', 'as', 'compared', 'to', 'the', 'national', 'average', 'of', '2.89', '%', 'the', 'official', 'said', ',', 'citing', 'the', '2019', 'Forest', 'Survey', 'of', 'India', 'report', '.', 'How', 'many', 'saplings', 'survive', '?', 'Uttar', 'Pradesh', 'State', 'Forest', 'Minister', 'Dara', 'Singh', 'said', 'the', 'long-term', 'survival', 'of', 'the', 'saplings', 'remains', 'a', 'concern', ',', 'adding', 'that', 'usually', 'only', '60', '%', 'of', 'the', 'saplings', 'survive', '.', 'The', 'rest', 'succumb', 'to', 'disease', 'or', 'lack', 'of', 'water', '.', 'However', ',', 'he', 'said', 'that', 'about', '80', '%', 'of', 'the', 'saplings', 'planted', 'in', 'the', 'last', 'four', 'annual', 'drives', 'have', 'survived', '.', 'All', 'the', 'regions', 'where', 'plantation', 'is', 'being', 'carried', 'out', 'have', 'been', 'geo-tagged', 'so', 'that', 'we', 'can', 'ascertain', 'what', 'exactly', 'happened', ',', 'Chauhan', 'told', 'The', 'Pioneer', 'newspaper', '.', 'These', 'saplings', 'carry', 'QR', 'codes', 'so', 'that', 'officials', 'can', 'maintain', 'a', 'record', 'and', 'verify', 'whether', 'the', 'saplings', 'survived', 'or', 'not', '.', 'Besides', ',', 'teams', 'have', 'been', 'formed', 'to', 'monitor', 'progress', 'of', 'the', 'plantation', 'drive', ',', 'he', 'said', '.', 'What', 'is', 'the', 'extent', 'of', 'India', "'s", 'tree', 'planting', 'project', '?', 'India', 'has', 'vowed', 'to', 'have', 'a', 'third', 'of', 'its', 'total', 'land', 'area', ',', 'or', '95', 'million', 'hectares', ',', 'under', 'forest', 'and', 'tree', 'cover', 'by', '2030', '.', 'The', 'government', 'has', 'allocated', '\$', '6.2', 'billion', '(', '€5.2', 'billion', ')', 'for', 'the', 'tree-planting', 'across', 'the', 'country', '.', 'However', ',', 'in', 'industrial', 'development', 'and', 'a', 'rapidly', 'growing', 'population', 'has', 'put', 'further', 'stress', 'on', 'the', 'land', '.'] ['Between', '2016', 'and', '2019', ',', 'the', 'state', 'forest', 'department', 'under', 'the', 'BJP', 'government', 'had', 'launched', 'Green', 'Maharashtra', 'drive', 'with', 'an', 'aim', 'to', 'plant', '50', 'crore', 'trees', 'across', 'the', 'state', 'in', 'the', 'four-year', 'period', '.', 'In', 'October', '2019', ',', 'the', 'government', 'had', 'claimed', 'it', 'had', 'surpassed', 'the', 'target', 'by', 'planting', '33', 'crore', 'trees', 'in', 'July-September', '2019', '.', 'The', 'Indian', 'Express', 'had', 'found', 'that', 'non-forest', 'agencies', '—', 'such', 'as', 'gram', 'panchayats', '—', 'which', 'were', 'tasked', 'with', 'planting', 'trees', 'had', 'not', 'uploaded', 'the', 'mandatory', 'audio-visual', 'proof', 'of', 'the', 'tree', 'plantation', 'drives', 'on', 'the', 'special

```

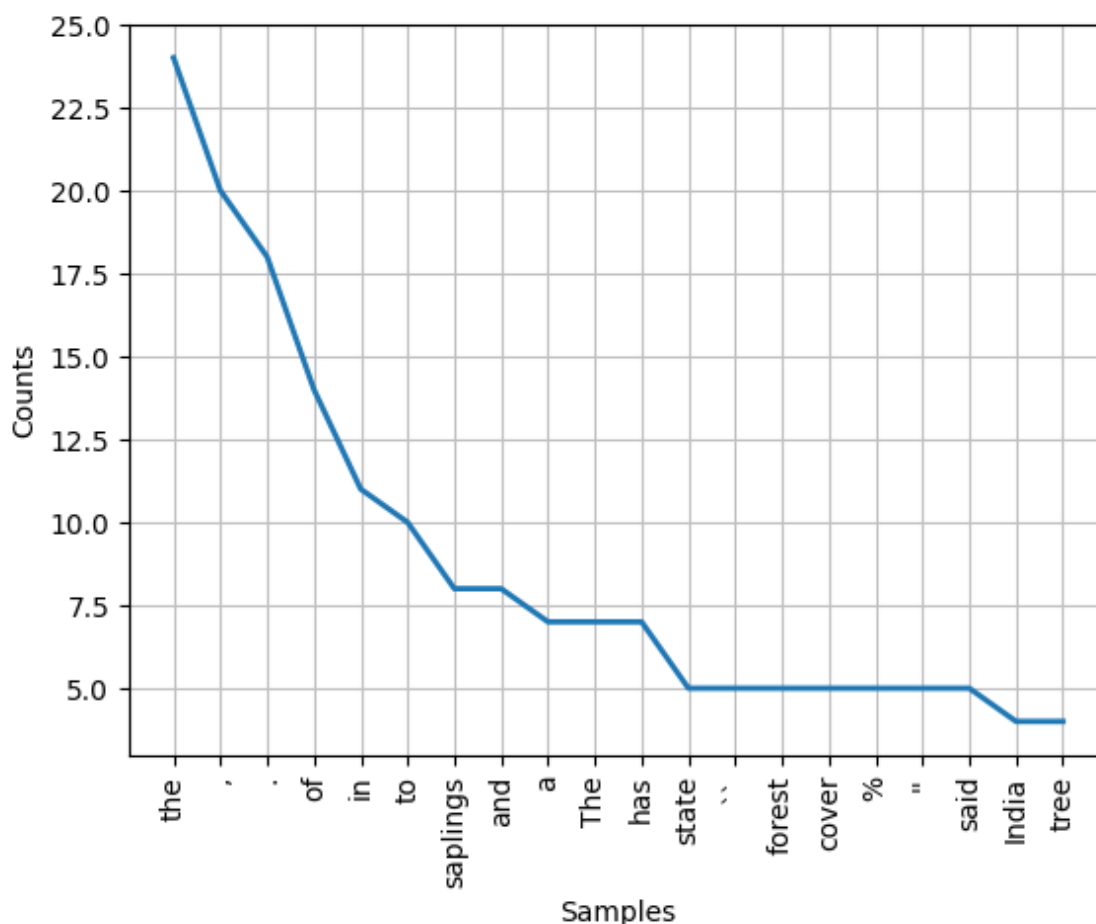
ly', 'created', 'portal', '.', 'In', 'Pune', 'Revenue', 'Division', ',',
'it', 'was', 'claimed', 'the', 'gram', 'panchayats', 'planted', '1.7', 'c
rore', 'saplings', ';', 'however', ',', 'no', 'evidence', 'was', 'uploade
d', 'for', '87', 'per', 'cent', '(', '1.49', 'crore', ')', 'saplings',
',', 'Also', ',', 'out', 'of', 'the', '59', 'government', 'agencies', 'in
volved', 'in', 'the', 'drive', 'as', 'many', 'as', '38', 'had', 'not', 's
ubmitted', 'survival', 'reports', 'about', 'the', 'saplings', '.', 'Thi
s', 'year', ',', 'the', 'targets', 'set', 'by', 'the', 'forest', 'departm
ent', 'were', 'comparatively', 'modest', '.', 'For', 'example', ',', 'Pun
e', 'Circle', '-', 'which', 'comprises', 'three', 'divisions', 'in', 'Pun
e', 'and', 'Solapur', 'district', '-', 'had', 'planned', 'to', 'plant',
In [20]: 17, 'lakh', 'saplings', 'on', 'forest', 'land', '.', 'However', ',', 'i
#to mark frequency distribution of words
from nltk.probability import FreqDist
word1 = FreqDist(word1)
freq1 = word1.most_common(1)
freq2 = word1.most_common(2)
print(freq1)
print(freq2)
', 'Daund', ',', 'Indapur', ',', 'Baramati', 'and', 'Havveli',
', 'plantations', 'were', 'done', 'for', 'plantation', 'of', 'about',
', '4', 'lakh', 'trees', 'with', 'special', 'emphasis', 'on', 'teakwood',
',', 'The', 'National', 'Forest', 'Policy', 'aims', 'and', 'emphasizes',
', 'to', 'maintain', '208', 'samples', 'and', '434', 'outcomes', 'country', ',', 's', 'geogra
phy', 'with', '296', 'samples', 'and', '678', 'outcomes', 'green', 'cover', '.', 'In',
', 'view', 'of', 'this', 'and', 'as', 'a', 'part', 'of', 'the', '50', 'crore
', 'plantation', 'programme', 'within', 'Maharashtra', ',', 'the', 'Maha
rashtra', 'Forest', 'Department', 'aims', 'to', 'plant', '4', 'crore', 's
aplings', 'all', 'over', 'the', 'State', 'between', '1st', 'to', 'July',
In [21]: 7th', 'most', 'common', 'to', 'celebrate', ',', 'Vanmohotsav', ',', 'The
e', 'plantation', 'programme', ',', 'which', 'was', 'announced', 'in', '2
016', 'Sunday', '1', 'the', 'aim', 'of', 'planting', '2', 'crore', 'trees', 'o
n', '1st', 'July', '2016', 'was', 'a', 'resounding', 'success', 'with',
', 'the', 'final', 'total', 'reported', 'figure', 'of', '2.82', 'crore', 'sa
plings', 'planted', 'on', 'a', 'single', 'day', '.', 'To', 'maintain', 'c
onsistency', 'of', 'this', 'platform', 'without', 'affecting', 'its', 'mo
mentum', 'the', 'Forest', 'Department', 'has', 'set', 'the', 'targe
t', 'of', '4', 'crore', ',', '13', 'crore', 'and', '3
3', 'campains', 'saplings', 'under', 'the', 'mission', 'of', '50', 'crore',
', 'plantation', 'which', 'shall', 'be', 'accomplished', 'in', 'the', 'thre
e', 'consecutive', 'years', 'viz', '.', '2017', ',', '2018', 'and', '201
9', 'The', '4', 'crore', 'saplings', 'for', 'the', 'year', '2017',
', 'will', 'be', 'planted', 'during', 'the', 'Vanmohotsav', ',', 'July', '1s
t', 'to', 'July', '7th', 'in', 'a', 'state-wide', 'drive', 'with', 'the',
', 'involvement', 'of', '33', 'government', 'departments', 'along', 'with',
', 'institutions', 'of', 'Schools', 'and', 'Colleges', ',', 'NSS', ',', 'NCC',
', 'CSR', 'NGOs', ',', 'Railways', ',', 'National', 'Highways',
', 'Defence', ',', 'NABARD', 'and', 'other', 'stakeholders', 'of', 'Soc
iety', 'change', 'in', 'a', 'first', 'of', 'its', 'kind', ',', 'a', '24-hour',
', 'toll-free', 'helpline', 'number', '1926', 'called', ',', 'Hello', 'Fo
rest', 'volunteers', 'has', 'been', 'set', 'up', 'to', 'provide', 'information',
', 'regarding', 'plantation', ',', 'protection', 'and', 'for', 'mass', 'awar
eness', '.', 'The', 'Forest', 'Department', 'has', 'created', 'a', 'mobil
e', 'application', 'called', ',', 'My', 'Plants', ',', 'to', 'record', 'd
etails', 'of', 'the', 'plantation', 'such', 'as', 'numbers', ',', 'specie
s', 'and', 'location', 'into', 'the', 'computer', 'system', 'of', 'the',
', 'Forest', 'Department', '.', 'All', 'volunteers', 'at', 'individual',
', 'collective', 'and', 'organizational', 'level', 'should', 'downloa
d', 'and', 'use', 'this', 'application', 'to', 'record', 'their', 'tree',
', 'plantation', 'work', 'through', 'the', 'application', ',', 'which', 'wil
l', 'be', 'operational', 'from', '1st', 'July', 'to', '7th', 'July', '.',
', 'In', 'consonance', 'of', 'the', 'public', 'participation', ',', 'the',
', 'Maharashtra', 'Forest', 'Department', 'has', 'initiated', 'the', ',', 'M
aharashtra', 'Harit', 'Sena', ',', 'Green', 'Army', ',', 'whic

```

```
h', 'is', 'a', 'body', 'of', 'dedicated', 'volunteers', 'to', 'participat
e', 'in', 'the', 'plantation', ',', 'protection', ',', 'and', 'activitie
s', 'in', 'forest', ',', 'wildlife', ',', 'and', 'related', 'sectors', 'a
round', 'the', 'year', '.', 'Individuals', 'and', 'organisations', 'inter
ested', 'in', 'volunteering', 'can', 'register', 'on', 'the', 'Green', 'A
rmy', 'website', 'www.greenarmy.mahaforest.gov.in', 'An', 'integrated',
'drive', 'has', 'been', 'set', 'in', 'place', 'to', 'ensure', 'seamless',
[and the successful', 'participation', 'from', 'all', 'stakeholders', 'o
f', 'society', ',', 'especially', 'the', 'public', '.']
('of', 22),
('.', 21),
('and', 18),
('to', 13),
('crore', 13),
('in', 11),
('plantation', 10),
('saplings', 9),
('a', 8),
('had', 7),
('\'', 7),
('with', 7),
('which', 7),
('on', 7),
('Forest', 7),
```

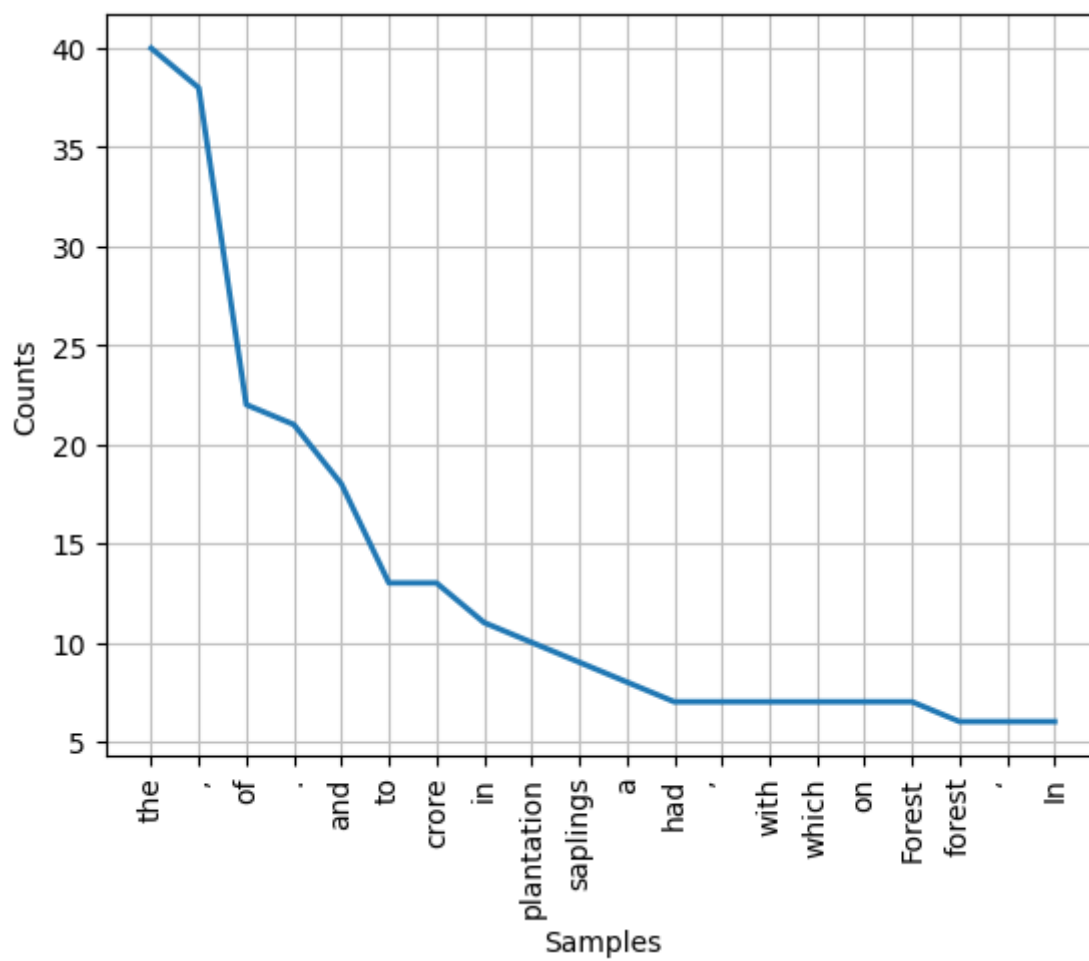
In [25]:

```
import matplotlib.pyplot as plt
freq1.plot(20)
plt.show()
```



In [34]:

```
freq2.plot(20)  
plt.show()
```



In [37]:



```
#now removing the stop words
from nltk.corpus import stopwords
stop_words = set(stopwords.words("english"))
print (stop_words)
```

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



In [39]:

```
#remove the stopwords
```

```
new_word1 = []
for i in word1:
    if i not in stop_words:
        new_word1.append(i)
```

```
print(new_word1)
```

```
['Millions', 'people', 'India', 'took', 'part', 'annual', 'tree', 'planti
ng', 'drive', 'Sunday', '.', 'More', '250', 'million', 'saplings', 'plant
ed', 'single', 'day', 'across', 'country', '"s', 'most-populous', 'stat
e', '.', 'The', 'campaign', 'led', 'Uttar', 'Pradesh', 'state', 'governme
nt', 'officials', ',', 'lawmakers', ',', 'activists', ',', 'bid', 'reduc
e', 'carbon', 'emissions', 'combat', 'climate', 'change', '.', 'Where',
'trees', 'planted', '?', 'The', 'saplings', 'planted', 'volunteers', 'for
ests', ',', 'farms', ',', 'schools', ',', 'along', 'riverbanks', 'highway
s', '.', '``', 'We', 'committed', 'increasing', 'forest', 'cover', 'Utta
r', 'Pradesh', '15', '%', 'total', 'land', 'area', 'next', 'five', 'year
s', ',', '"', 'said', 'state', 'forest', 'official', 'Manoj', 'Singh',
',', 'According', 'another', 'government', 'official', ',', 'forest', 'co
ver', 'state', 'increased', 'last', 'years', '.', '``', 'There', 'increas
e', '127', 'sqare', 'kilometers', '(', '79', 'sqare', 'miles', ')', 'fore
st', 'cover', 'Uttar', 'Pradesh', 'compared', '2017', ',', '"', 'state',
'government', 'spokesperson', 'quoted', 'saying', 'The', 'Indian', 'Expre
ss', 'newspaper', '.', '``', 'There', 'also', 'increase', 'trees', 'plant
s', '.', 'The', 'tree', 'cover', 'increased', '3.05', '%', ',', 'compare
d', 'national', 'average', '2.89', '%', ',', '"', 'official', 'said',
',', 'citing', '2019', 'Forest', 'Survey', 'India', 'report', '.', 'How',
'many', 'saplings', 'survive', '?', 'Uttar', 'Pradesh', 'State', 'Fores
t', 'Minister', 'Dara', 'Singh', 'said', 'long-term', 'survival', 'saplin
gs', 'remains', 'concern', ',', 'adding', 'usually', '60', '%', 'sapling
s', 'survive', '.', 'The', 'rest', 'succumb', 'disease', 'lack', 'water',
',', 'However', ',', 'said', '80', '%', 'saplings', 'planted', 'last', 'f
our', 'annual', 'drives', 'survived', '.', '``', 'All', 'regions', 'plant
ation', 'carried', 'geo-tagged', 'ascertain', 'exactly', 'happened', ',',
'', 'Chauhan', 'told', 'The', 'Pioneer', 'newspaper', '.', '``', 'Thes
e', 'saplings', 'carry', 'QR', 'codes', 'officials', 'maintain', 'recor
d', 'verify', 'whether', 'saplings', 'survived', '.', 'Besides', ',', 'te
ams', 'formed', 'monitor', 'progress', 'plantation', 'drive', ',', '"',
'said', '.', 'What', 'extent', 'India', '"s', 'tree', 'planting', 'projec
t', '?', 'India', 'vowed', 'third', 'total', 'land', 'area', ',', '95',
'million', 'hectares', ',', 'forest', 'tree', 'cover', '2030', '.', 'Th
e', 'government', 'allocated', '$', '6.2', 'billion', '(', '€5.2', 'billi
on', ')', 'tree-planting', 'across', 'country', '.', 'However', ',', 'ind
ustrial', 'development', 'rapidly', 'growing', 'population', 'put', 'stre
ss', 'land', '.']
```


In [41]:



```
new_word2 = []  
for i in word2:  
    if i not in stop_words:  
        new_word2.append(i)  
  
print(new_word2)
```



```
[ 'Between', '2016', '2019', ',', 'state', 'forest', 'department', 'BJP',
'government', 'launched', ',', 'Green', 'Maharashtra', ',', 'drive', 'ai
m', 'plant', '50', 'crore', 'trees', 'across', 'state', 'four-year', 'per
iod', 'In', 'October', '2019', ',', 'government', 'claimed', 'surpas
sed', 'target', 'planting', '33', 'crore', 'trees', 'July-September', '20
19', 'The', 'Indian', 'Express', 'found', 'non-forest', 'agencies',
',', 'gram', 'panchayats', '-', 'tasked', 'planting', 'trees', 'uploade
d', 'mandatory', 'audio-visual', 'proof', 'tree', 'plantation', 'drives',
'specially', 'created', 'portal', '.', 'In', 'Pune', 'Revenue', 'Divisio
n', ',', 'claimed', 'gram', 'panchayats', 'planted', '1.7', 'crore', 'sap
lings', 'However', 'Porter', 'Stemmer', 'uploaded', '87', 'per', 'cent',
ps = PorterStemmer(), 'saplings', '.', 'Also', ',', '59', 'governmen
t', 'agencies', 'involved', 'drive', 'many', '38', 'submitted', 'surviva
l', 'reports', 'saplings', '.', 'This', 'year', ',', 'targets', 'set', 'f
or', 'department', 'comparatively', 'modest', '.', 'For', 'example',
',', 'Pune', 'Circle', '-', 'comprises', 'three', 'divisions', 'Pune', 'S
atapur', 'district', '-', 'planned', 'plant', '17', 'lakh', 'saplings',
for forest in new word1: '.', 'However', ',', 'may', 'meet', 'target', 'due', 'u
navability', 'and', 'ps:stem(1st', 'year', 'Pune', 'Circle', 'plante
print("Old words")', 'saplings', 'forest', 'land', '.', 'In', 'Pune', 'Divis
ion', 'comprises', 'six', 'talukas', 'namely', 'Maval', ',', 'Mulsh
print("\n\nNew Stemmed words")', 'Baramati', 'Havveli', '-', 'prepa
rate(stem_words1)', 'plantation', '4', 'lakh', 'trees', 'special', 'emphasi
s', 'teakwood', '.', 'The', 'National', 'Forest', 'Policy', 'aims', 'emph
asizes', 'maintaining', '33', '%', 'country', ',', 'geographical', 'are
Old words', 'green', 'cover', '.', 'In', 'view', 'part', '50', 'crore',
'plantations', 'programme', 'India', 'within', 'Maharashtra', 'annual', 'MahafashtPa
n', 'Forest', 'Department', 'aims', 'plant', '250', 'million', 'saplings', 'stat
e', 'planted', 'single', '7th', 'day', 'across', 'country', 's', 'most', 'vanmohitsav',
s', 'state', 'The', 'plantation', 'campaign', 'led', 'Uttar', 'announced', '2016', 'ai
m', 'planting', '2', 'officials', 'trees', 'lawmakers', 'July', '2016', 'resoundin
g', 'success', 'reducing', 'carbon', 'emissions', 'combat', 'climate', 'crisis',
saplings', 'where', 'trees', 'single', 'planted', '1', 'The', 'maintains', 'consisten
cy', 'volunteers', 'with', 'forest', 'affecting', 'momentum', 'schools', 'Forest', 'alop', 'Depa
rtment', 'riverbanks', 'highways', 'plantation', '4', 'We', 'committed', '13', 'increasing',
833', 'forest', 'cover', 'Uttar', 'Mississipp', '50', '15', 'crore%', 'plantation', 'land', 'shal
l', 'area', 'accomplished', 'five', 'three', 'years', 'consecutive', 'said', 'state', 'forest', '2017',
'official', '2018', '2019', 'Singh', 'The', '4', 'According', 'saplings', 'government', '201
7', 'official', 'vanmohitsav', 'cover', 'July', 'state', 'increased', '7th', 'last', 'state-w
ide', 'years', 'drive', 'involvement', 'increase', 'government', 'square', 'kilometers', 'along',
'70', 'schools', 'miles', 'colleges', 'forest', 'NSS', 'cover', 'Nuttari', 'Prades
h', 'compared', '2017', 'Railways', 'National', 'government', 'spokesperson', 'Ac
e', 'quoted', 'NABARD', 'paying', 'stakeholders', 'Indian', 'Express', 'in', 'news', 'part', 'kind',
',', '24', 'hour', 'also', 'increase', 'helpline', 'Number', '1926', 'called',
',', 'Hello', 'Forest', ',', 'set', 'provide', 'information', 'regardin
g', 'plantation', ',', 'protection', 'mass', 'awareness', '.', 'The', 'Fo
rest', 'Department', 'created', 'mobile', 'application', 'called', ',',
'My', 'Plants', ',', 'record', 'details', 'plantation', 'numbers', ',',
'species', 'location', 'computer', 'system', 'Forest', 'Department', '.',
'All', 'volunteers', 'individual', ',', 'collective', 'organizational',
'level', 'download', 'use', 'application', 'record', 'tree', 'plantatio
n', 'work', 'application', ',', 'operational', '1st', 'July', '7th', 'Jul
y', '.', 'In', 'consonance', 'public', 'participation', ',', 'Maharashtr
a', 'Forest', 'Department', 'initiated', ',', 'Maharashtra', 'Harit', 'Se
na', ',', '/', ',', 'Green', 'Army', ',', 'body', 'dedicated', 'volunteer
s', 'participate', 'plantation', ',', 'protection', ',', 'activities', 'f
orest', ',', 'wildlife', ',', 'related', 'sectors', 'around', 'year',
',', 'Individuals', 'organisations', 'interested', 'volunteering', 'regis
ter', 'Green', 'Army', 'website', 'www.greenarmy.mahaforest.gov.in', 'A
n', 'integrated', 'drive', 'set', 'place', 'ensure', 'seamless', 'success
```

```
ful', 'participation', 'stakeholders', 'society', ',', 'especially', 'pub  
lic[47]:']
```



```
stem_words2 = []  
for i in new_word2:  
    stem_words2.append(ps.stem(i))  
print("Old words")  
print(new_word2)  
print("\n\n\nNew Stemmed words")  
print(stem_words2)
```


Old words

['Between', '2016', '2019', ',', 'state', 'forest', 'department', 'BJP', 'government', 'launched', '(', 'Green', 'Maharashtra', ')', 'drive', 'aim', 'plant', '50', 'crore', 'trees', 'across', 'state', 'four-year', 'period', '.', 'In', 'October', '2019', ',', 'government', 'claimed', 'surpassed', 'target', 'planting', '33', 'crore', 'trees', 'July-September', '2019', '.', 'The', 'Indian', 'Express', 'found', 'non-forest', 'agencies', '-', 'gram', 'panchayats', '-', 'tasked', 'planting', 'trees', 'uploaded', 'mandatory', 'audio-visual', 'proof', 'tree', 'plantation', 'drives', 'specially', 'created', 'portal', '.', 'In', 'Pune', 'Revenue', 'Division', ',', 'claimed', 'gram', 'panchayats', 'planted', '1.7', 'crore', 'saplings', ';', 'however', ',', 'evidence', 'uploaded', '87', 'per', 'cent', '(', '1.49', 'crore', ')', 'saplings', '.', 'Also', ',', '59', 'government', 'agencies', 'involved', 'drive', 'many', '38', 'submitted', 'survival', 'reports', 'saplings', '.', 'This', 'year', ',', 'targets', 'set', 'forest', 'department', 'comparatively', 'modest', '.', 'For', 'example', ',', 'Pune', 'Circle', '-', 'comprises', 'three', 'divisions', 'Pune', 'Solapur', 'district', '-', 'planned', 'plant', '17', 'lakh', 'saplings', 'forest', 'land', '.', 'However', ',', 'may', 'meet', 'target', 'due', 'unavailability', 'funds', '.', 'Last', 'year', 'Pune', 'Circle', 'planted', '70', 'lakh', 'saplings', 'forest', 'land', '.', 'In', 'Pune', 'Division', '-', 'comprises', 'six', 'talukas', 'namely', 'Maval', ',', 'Mulshi', ',', 'Daund', ',', 'Indapur', ',', 'Baramati', 'Havveli', '-', 'preparations', 'done', 'plantation', '4', 'lakh', 'trees', 'special', 'emphasis', 'teakwood', '.', 'The', 'National', 'Forest', 'Policy', 'aims', 'emphasizes', 'maintaining', '33', '%', 'country', 'geographical', 'area', 'forest', 'green', 'cover', '.', 'In', 'view', 'part', '50', 'crore', 'plantation', 'programme', 'within', 'Maharashtra', ',', 'Maharashtra', 'Forest', 'Department', 'aims', 'plant', '4', 'crore', 'saplings', 'State', '1st', 'July', '7th', ',', '2017', 'celebrate', '(', 'Vanmohotsav', ')', '.', 'The', 'plantation', 'programme', ',', 'announced', '2016', 'aim', 'planting', '2', 'crore', 'trees', '1st', 'July', '2016', 'resounding', 'success', 'final', 'total', 'reported', 'figure', '2.82', 'crore', 'saplings', 'planted', 'single', 'day', '.', 'To', 'maintain', 'consistency', 'platform', 'without', 'affecting', 'momentum', ',', 'Forest', 'Department', 'set', 'target', 'plantation', '4', 'crore', ',', '13', 'crore', '33', 'crore', 'saplings', 'mission', '50', 'crore', 'plantation', 'shall', 'accomplished', 'three', 'consecutive', 'years', 'viz', '.', '2017', ',', '2018', '2019', '.', 'The', '4', 'crore', 'saplings', 'year', '2017', 'planted', 'Vanmohotsav', ',', 'July', '1st', 'July', '7th', 'state-wide', 'drive', 'involvement', '33', 'government', 'departments', 'along', 'Students', 'Schools', 'Colleges', ',', 'NSS', ',', 'NCC', ',', 'CSR', ',', 'NGOs', ',', 'Railways', ',', 'National', 'Highways', ',', 'Defence', ',', 'NABARD', 'stakeholders', 'Society', '.', 'In', 'first', 'kind', ',', '24-hour', 'toll', 'free', 'helpline', 'number', '1926', 'called', '(', 'Hello', 'Forest', ')', 'set', 'provide', 'information', 'regarding', 'plantation', ',', 'protection', 'mass', 'awareness', '.', 'The', 'Forest', 'Department', 'created', 'mobile', 'application', 'called', '(', 'My', 'Plants', ')', 'record', 'details', 'plantation', 'numbers', ',', 'species', 'location', 'computer', 'system', 'Forest', 'Department', '.', 'All', 'volunteers', 'individual', ',', 'collective', 'organizational', 'level', 'download', 'use', 'application', 'record', 'tree', 'plantation', 'work', 'application', ',', 'operational', '1st', 'July', '7th', 'July', '.', 'In', 'consonance', 'public', 'participation', ',', 'Maharashtra', 'Forest', 'Department', 'initiated', '(', 'Maharashtra', 'Harit', 'Sena', ')', 'Green', 'Army', 'body', 'dedicated', 'volunteers', 'participate', 'plantation', ',', 'protection', ',', 'activities', 'forest', ',', 'wildlife', ',', 'related', 'sectors', 'around', 'year', '.', 'Individuals', 'organisations', 'interested', 'volunteering', 'register', 'Green', 'Army', 'website', 'www.greenarmy.mahaforest.gov.in', 'An', 'integrated', 'drive', 'set', 'place', 'ensure', 'seamless', 'success

```
ful', 'participation', 'stakeholders', 'society', ',', 'especially', 'public', '.']
```

```
In [49]:
```

```
#Lemmatization
```

```
#converting word in its base form like better converted to good
```

```
New Stemmed words
```

```
['between', '2016', '2019', ',', 'state', 'forest', 'depart', 'bjp', 'gov',
'ann[52]launch', ',', 'green', 'maharashtra', ',', 'drive', 'aim', 'plan',
t', '50', 'crore', 'tree', 'across', 'state', 'four-year', 'period', '.',
'from', 'nltk.stem.wordnet import WordNetLemmatizer', 'surpass', 'target', 'plan',
wnl = WordNetLemmatizer(), 'july-septemb', '2019', '.', 'the', 'indian',
'express', 'found', 'non-forest', 'agenc', '-', 'gram', 'panchayat', '-',
'task', 'plant', 'tree', 'upload', 'mandatori', 'audio-visu', 'proof', 't',
'ee[55]plantat', 'drive', 'special', 'creat', 'portal', '.', 'in', 'pun',
e', 'revenu', 'divis', ',', 'claim', 'gram', 'panchayat', 'plant', '1.7',
lem_words1 = [], 'howev', ',', 'evid', 'upload', '87', 'per', 'cen',
for i in new_word1: 'sapl', '.', 'also', ',', '59', 'govern',
'agenc', 'involv', 'drive', 'mahat', 'submit', 'surviv', 'report',
print("Old words")
sapl.append(thi, 'year', ',', 'target', 'set', 'forest', 'depart', 'co',
mpar(new_word1), 'for', 'exampl', ',', 'pune', 'circl', '-', 'compri',
print("\n\nNew Lemmatized words")
print(three, divis, pune, solapur, 'district', '-', 'plan', 'plan',
print(lem_words1) 'sapl', 'forest', 'land', '.', 'howev', ',', 'may', 'me',
et', 'target', 'due', 'unavail', 'fund', '.', 'last', 'year', 'pune', 'ci',
rcl', 'plant', '70', 'lakh', 'sapl', 'forest', 'land', '.', 'in', 'pune',
Old words: 'compris', 'six', 'taluka', 'name', 'maval', ',', 'mulshi',
[ 'Millions', 'people', 'india', 'took', 'part', 'annual', 'tree', 'pl',
ning', 'drive', 'sunday', 'More', '250', 'million', 'saplings', 'do',
ne', 'plantat', '4', 'lakh', 'tree', 'special', 'emphas', 'teakwood',
'planted', 'single', 'day', 'across', 'country', 'S', 'most', 'populou', '3',
s', 'state', 'countri', 'The', 'campaign', 'led', 'Uttar', 'Pradesh', 'stat',
e', 'government', 'officials', 'geograph', 'area', 'forest', 'green', 'cover',
n', 'bid', 'reduce', 'part', 'carbon', 'emissions', 'combat', 'climate', 'cha',
nge', '4', 'crore', 'where', 'trees', 'planted', '1st', 'juli', 'saplings', 'plant', 'ce',
ed', 'volunteers', 'Vanmohot', 's', 'farm', 'plantat', 'schools', 'program', 'alon',
g', 'riverbanks', 'highways', 'plant', '2', 'crore', 'committed', '1st', 'juli',
g', 'forest', 'resound', 'success', 'final', 'total', 'report', 'figur', '2:8',
2', 'crore', 'next', 'five', 'years', 'singl', 'day', 'said', 'state', 'forest',
2', 'official', 'Manoj', 'Singh', 'According', 'another', 'governmen',
sist', 'platform', 'without', 'affect', 'momentum', 'forest', 'depar',
t', 'official', 'target', 'plantat', '4', 'crore', 'state', 'increased', 'last', '33',
'crore', 'sapl', 'mission', '50', 'crore', '127', 'square', 'shall', 'accomplis',
h', 'three', 'square', 'miles', 'year', 'forest', 'cover', 'Uttar', 'Prades', '2019',
n', 'compared', '4', 'crore', 'sapl', 'year', 'state', '2017', 'governmen', 'spokesperso',
n', 'quoted', '1st', 'juli', 'The', '7th', 'Indian', 'State', 'Express', 'newspaper', 's',
'govern', 'There', 'also', 'along', 'increase', 'trees', 'school', 'plant', 'involv', '33',
',', 'ncc', ',', 'csr', ',', 'ngo', ',', 'railway', ',', 'nation', 'highw',
ay', ',', 'defenc', ',', 'nabard', 'stakehold', 'societi', '.', 'in', 'fi',
rst', 'kind', ',', '24-hour', 'toll', 'free', 'helplin', 'number', '192',
6', 'call', ',', 'hello', 'forest', ',', 'set', 'provid', 'inform', 'rega',
rd', 'plantat', ',', 'protect', 'mass', 'awar', '.', 'the', 'forest', 'de',
part', 'creat', 'mobil', 'applic', 'call', ',', 'my', 'plant', ',', 'reco',
rd', 'detail', 'plantat', 'number', ',', 'speci', 'locat', 'comput', 'sys',
tem', 'forest', 'depart', '.', 'all', 'volunt', 'individu', ',', 'collec',
t', 'organiz', 'level', 'download', 'use', 'applic', 'record', 'tree', 'p',
lantat', 'work', 'applic', ',', 'oper', '1st', 'juli', '7th', 'juli',
',', 'in', 'conson', 'public', 'particip', ',', 'maharashtra', 'forest',
'depart', 'initi', ',', 'maharashtra', 'harit', 'sena', ',', '/', ',', 'g',
reen', 'armi', ',', 'bodi', 'dedic', 'volunt', 'particip', 'plantat',
',', 'protect', ',', 'activ', 'forest', ',', 'wildlif', ',', 'relat', 'se',
ctor', 'around', 'year', '.', 'individu', 'organis', 'interest', 'volun',
t', 'regist', 'green', 'armi', 'websit', 'www.greenarmy.mahaforest.gov.i
```

n', 'an', 'integr', 'drive', 'set', 'place', 'ensur', 'seamless', 'succes
s', 'particip', 'stakehold', 'societi', ',', 'especi', 'public', '.']



```
lem_words2 = []  
for i in new_word2:  
    lem_words2.append(wnl.lemmatize(i))  
print("Old Words")  
print(new_word1)  
print("\n\n\nNew Lemmatized words")  
print(lem_words2)
```


Old Words

['Millions', 'people', 'India', 'took', 'part', 'annual', 'tree', 'planting', 'drive', 'Sunday', '.', 'More', '250', 'million', 'saplings', 'planted', 'single', 'day', 'across', 'country', "'s", 'most-populous', 'state', '.', 'The', 'campaign', 'led', 'Uttar', 'Pradesh', 'state', 'government', 'officials', ',', 'lawmakers', ',', 'activists', ',', 'bid', 'reduce', 'carbon', 'emissions', 'combat', 'climate', 'change', '.', 'Where', 'trees', 'planted', '?', 'The', 'saplings', 'planted', 'volunteers', 'forests', ',', 'farms', ',', 'schools', ',', 'along', 'riverbanks', 'highways', '.', 'We', 'committed', 'increasing', 'forest', 'cover', 'Uttar', 'Pradesh', '15', '%', 'total', 'land', 'area', 'next', 'five', 'years', ',', 'said', 'state', 'forest', 'official', 'Manoj', 'Singh', '.', 'According', 'another', 'government', 'official', ',', 'forest', 'cover', 'state', 'increased', 'last', 'years', '.', 'There', 'increase', '127', 'square', 'kilometers', '["', '79', 'square', 'miles', '"]', 'forest', 'cover', 'Uttar', 'Pradesh', 'compared', '2017', ',', 'state', 'government', 'spokesperson', 'quoted', 'saying', 'The', 'Indian', 'Express', 'newspaper', '.', 'There', 'also', 'increase', 'trees', 'plants', '.', 'The', 'tree', 'cover', 'increased', '3.05', '%', ',', 'compared', 'national', 'average', '2.89', '%', ',', 'official', 'said', ',', 'citing', '2019', 'Forest', 'Survey', 'India', 'report', '.', 'How', 'many', 'saplings', 'survive', '?', 'Uttar', 'Pradesh', 'State', 'Forest', 'Minister', 'Dara', 'Singh', 'said', 'long-term', 'survival', 'saplings', 'remains', 'concern', ',', 'adding', 'usually', '60', '%', 'saplings', 'survive', '.', 'The', 'rest', 'succumb', 'disease', 'lack', 'water', '.', 'However', ',', 'said', '80', '%', 'saplings', 'planted', 'last', 'four', 'annual', 'drives', 'survived', '.', 'All', 'regions', 'plantation', 'carried', 'geo-tagged', 'ascertain', 'exactly', 'happened', ',', 'Chauhan', 'told', 'The', 'Pioneer', 'newspaper', '.', 'These', 'saplings', 'carry', 'QR', 'codes', 'officials', 'maintain', 'record', 'verify', 'whether', 'saplings', 'survived', '.', 'Besides', ',', 'teams', 'formed', 'monitor', 'progress', 'plantation', 'drive', ',', 'said', '.', 'What', 'extent', 'India', "'s", 'tree', 'planting', 'project', '?', 'India', 'vowed', 'third', 'total', 'land', 'area', ',', '95', 'million', 'hectares', ',', 'forest', 'tree', 'cover', '2030', '.', 'The', 'government', 'allocated', '\$', '6.2', 'billion', '(', '€5.2', 'billion', ')', 'tree-planting', 'across', 'country', '.', 'However', ',', 'industrial', 'development', 'rapidly', 'growing', 'population', 'put', 'stress', 'land', '.', 'Between', '2016', '2019', ',', 'state', 'forest', 'department', 'BJP', 'government', 'launched', 'Green', 'Maharashtra', 'drive', 'aim', 'plant', '50', 'crore', 'trees', 'across', 'state', 'four-year', 'period', '.', 'In', 'October', '2019', ',', 'government', 'claimed', 'surpassed', 'target', 'planting', '33', 'crore', 'trees', 'July-September', '2019', '.', 'The', 'Indian', 'Express', 'found', 'non-forest', 'agencies', '-', 'gram', 'panchayats', '-', 'tasked', 'planting', 'trees', 'uploaded', 'mandatory', 'audio-visual', 'proof', 'tree', 'plantation', 'drives', 'specially', 'created', 'portal', '.', 'In', 'Pune', 'Revenue', 'Division', ',', 'claimed', 'gram', 'panchayats', 'planted', '1.7', 'crore', 'saplings', ';', 'however', ',', 'evidence', 'uploaded', '87', 'per', 'cent', '(', '1.49', 'crore', ')', 'saplings', '.', 'Also', ',', '59', 'government', 'agencies', 'involved', 'drive', 'many', '38', 'submitted', 'survival', 'reports', 'saplings', '.', 'This', 'year', ',', 'targets', 'set', 'forest', 'department', 'comparatively', 'modest', '.', 'For', 'example', ',', 'Pune', 'Circle', '-', 'comprises', 'three', 'divisions', 'Pune', 'Solapur', 'district', '-', 'planned', 'plant', '17', 'lakh', 'saplings', 'forest', 'land', '.', 'However', ',', 'may', 'meet', 'target', 'due', 'unavailability', 'funds', '.', 'Last', 'year', 'Pune', 'Circle', 'planted', '70', 'lakh', 'saplings', 'forest', 'land', '.', 'In', 'Pune', 'Division', '-', 'comprises', 'six', 'talukas', 'namely', 'Maval', 'Mulshi', 'Daund', 'Indapur', 'Baramati', 'Haveli', '-', 'preparations', 'done', 'plantation', '4', 'lakh', 'trees', 'spe

cial', 'emphasis', 'teakwood', '.', 'The', 'National', 'Forest', 'Polic
y', 'aims', 'emphasizes', 'maintaining', '33', '%', 'country', '', 'geog
raphical', 'area', 'forest', 'green', 'cover', '.', 'In', 'view', 'part',
'50', 'crore', 'plantation', 'programme', 'within', 'Maharashtra', '',
'Maharashtra', 'Forest', 'Department', 'aims', 'plant', '4', 'crore', 'sa
plings', 'State', '1st', 'July', '7th', '', '2017', 'celebrate', '', 'V
anmohotsav', '', '.', 'The', 'plantation', 'programme', '', 'announ
ced', '2016', 'aim', 'planting', '2', 'crore', 'trees', '1st', 'July', '201
6', 'resounding', 'success', 'final', 'total', 'reported', 'figure', '2.8
2', 'crore', 'saplings', 'planted', 'single', 'day', '.', 'To', 'maintai
n', 'consistency', 'platform', 'without', 'affecting', 'momentum', '',
'Forest', 'Department', 'set', 'target', 'plantation', '4', 'crore', '',
'13', 'crore', '33', 'crore', 'saplings', 'mission', '50', 'crore', 'plan
tation', 'shall', 'accomplished', 'three', 'consecutive', 'years', 'viz',
, '.', '2017', '', '2018', '2019', '.', 'The', '4', 'crore', 'saplings',
'year', '2017', 'planted', 'Vanmohotsav', '', 'July', '1st', 'July', '7t
h', 'state-wide', 'drive', 'involvement', '33', 'government', 'departmen
t', 'along', 'Students', 'Schools', 'Colleges', '', 'NSS', '', 'NCC',
, '', 'CSR', '', 'NGOs', '', 'Railways', '', 'National', 'Highways',
, '', 'Defence', '', 'NABARD', 'stakeholders', 'Society', '.', 'In', 'fir
st', 'kind', '', '24-hour', 'toll', 'free', 'helpline', 'number', '192
6', 'called', '', 'Hello', 'Forest', '', 'set', 'provide', 'informatio
n', 'regarding', 'plantation', '', 'protection', 'mass', 'awareness',
, '.', 'The', 'Forest', 'Department', 'created', 'mobile', 'application',
'called', '', 'My', 'Plants', '', 'record', 'details', 'plantation', 'n
umbers', '', 'species', 'location', 'computer', 'system', 'Forest', 'Dep
artment', '.', 'All', 'volunteers', 'individual', '', 'collective', 'org
anizational', 'level', 'download', 'use', 'application', 'record', 'tre
e', 'plantation', 'work', 'application', '', 'operational', '1st', 'Jul
y', '7th', 'July', '.', 'In', 'consonance', 'public', 'participation',
, '', 'Maharashtra', 'Forest', 'Department', 'initiated', '', 'Maharashtr
a', 'Harit', 'Sena', '', '/', '', 'Green', 'Army', '', 'body', 'dedica
ted', 'volunteers', 'participate', 'plantation', '', 'protection', '',
'activities', 'forest', '', 'wildlife', '', 'related', 'sectors', 'arou
nd', 'year', '.', 'Individuals', 'organisations', 'interested', 'voluntee
ring', 'register', 'Green', 'Army', 'website', 'www.greenarmy.mahaforest.
gov.in', 'An', 'integrated', 'drive', 'set', 'place', 'ensure', 'seamles
s', 'successful', 'participation', 'stakeholders', 'society', '', 'espec
ially', 'public', '.']

New Lemmatized words

['Between', '2016', '2019', '', 'state', 'forest', 'department', 'BJP',
'government', 'launched', '', 'Green', 'Maharashtra', '', 'drive', 'ai
m', 'plant', '50', 'crore', 'tree', 'across', 'state', 'four-year', 'peri
od', 'In', 'October', '2019', 'government', 'claimed', 'surpass
ed', 'target', 'planting', '33', 'crore', 'tree', 'July-September', '201
9', '.', 'The', 'Indian', 'Express', 'found', 'non-forest', 'agency',
'-', 'gram', 'panchayat', '-', 'tasked', 'planting', 'tree', 'uploaded',
'mandatory', 'audio-visual', 'proof', 'tree', 'plantation', 'drive', 'spe
cially', 'created', 'portal', '.', 'In', 'Pune', 'Revenue', 'Division',
, '', 'claimed', 'gram', 'panchayat', 'planted', '1.7', 'crore', 'saplin
g', ';', 'however', '', 'evidence', 'uploaded', '87', 'per', 'cent',
'(', '1.49', 'crore', ')', 'sapling', '.', 'Also', '', '59', 'governmen
t', 'agency', 'involved', 'drive', 'many', '38', 'submitted', 'survival',
'report', 'sapling', '.', 'This', 'year', '', 'target', 'set', 'forest',
'department', 'comparatively', 'modest', '.', 'For', 'example', '', 'Pun
e', 'Circle', '-', 'comprises', 'three', 'division', 'Pune', 'Solapur',
'district', '-', 'planned', 'plant', '17', 'lakh', 'sapling', 'forest',



```

'land', '.', 'However', ',', 'may', 'meet', 'target', 'due', 'unavailabil
ity', 'fund', '.', 'Last', 'year', 'Pune', 'Circle', 'planted', '70', 'la
kh', 'sapling', 'forest', 'land', '.', 'In', 'Pune', 'Division', '-', 'co
mpprises six', 'talukas', 'namely', 'Maval', ',', 'Mulshi', ',', 'Daun
nitk, pos, tag(word1)
d', ',', 'Indapur', ',', 'Baramati', 'Havveli', '-', 'preparation', 'don
e', 'plantation', 'NNS4', 'lakh', 'tree', 'special', 'emphasis', 'teakwood',
'('in', 'The', 'IN', 'National', 'Forest', 'Policy', 'aim', 'emphasizes', 'maintai
ning forests', 'NNS', 'country', ',', 'geographical', 'area', 'forest', 'gree
n', 'cover', 'In', 'view', 'part', '50', 'crore', 'plantation', 'pro
gramme', 'within', 'Maharashtra', ',', 'Maharashtra', 'Forest', 'Departme
nt', 'aim', 'plant', '4', 'crore', 'sapling', 'State', '1st', 'July', '7t
h', 'schools', '2017', 'celebrate', '(', 'Vanmohotsav', ',', 'The', 'plant
ation', 'programme', 'announced', '2016', 'aim', 'planting', '2', 'c
rore', 'tree', '1st', 'July', '2016', 'resounding', 'success', 'final',
'total', 'reported', 'figure', '2.82', 'crore', 'sapling', 'planted', 'si
ngle riverbanks', 'NNS', 'maintain', 'consistency', 'platform', 'withou
t', 'affecting', 'momentum', ',', 'Forest', 'Department', 'set', 'targe
t', 'highways', 'NNS', '4', 'crore', ',', '13', 'crore', '33', 'crore', 'sapli
ng', 'mission', '50', 'crore', 'plantation', 'shall', 'accomplished', 'th
ree', 'consecutive', 'year', 'viz', '.', '2017', ',', '2018', '2019',
'('we', 'The', 'PRP', '4', 'crore', 'sapling', 'year', '2017', 'planted', 'Vanmohot
sav', 'are', 'VB', '1st', 'July', '7th', 'state-wide', 'drive', 'involvem
ent', 'committed', 'government', 'department', 'along', 'Students', 'Schools',
'Colleges', 'TO', 'NSS', ',', 'NCC', ',', 'CSR', ',', 'NGOs', ',', 'Railwa
ys', 'increasing', 'National', 'Highways', ',', 'Defence', ',', 'NABARD', 'stakeho
lder', 'Society', '.', 'In', 'first', 'kind', ',', '24-hour', 'toll', 'fr
ee', 'helpline', 'number', '1926', 'called', '(', 'Hello', 'Forest', ',',
'set', 'provide', 'information', 'regarding', 'plantation', ',', 'protect
ion', 'awareness', '.', 'The', 'Forest', 'Department', 'created',
'mobile', 'application', 'called', '(', 'My', 'Plants', ',', 'record', 'd
etail', 'state', 'plantation', 'number', ',', 'specie', 'location', 'computer', 's
ystem', 'Forest', 'Department', '.', 'All', 'volunteer', 'individual',
'('department', 'NNS', 'organizational', 'level', 'download', 'use', 'applica
tion', 'Under', 'record', 'tree', 'plantation', 'work', 'application', ',', 'opera
tional', '1st', 'July', '7th', 'July', '.', 'In', 'consonance', 'public',
'participation', ',', 'Maharashtra', 'Forest', 'Department', 'initiated',
'('government', 'NNS', 'Harit', 'Sena', ',', '/', '(', 'Green', 'Army', ',',
'body', 'dedicated', 'volunteer', 'participate', 'plantation', ',', 'prot
ection', 'activity', 'forest', ',', 'wildlife', ',', 'related', 'sec
tor', 'around', 'year', '.', 'Individuals', 'organisation', 'interested',
'volunteer', 'register', 'Green', 'Army', 'website', 'www.greenarmy.ma
hara.gov.in', 'NAP', 'integrated', 'drive', 'set', 'place', 'ensure',
'seamless', 'successful', 'participation', 'stakeholder', 'society', ',',
'especially', 'public', '.']
('with', 'IN'),
('an', 'DT'),
('aim', 'NN'),
('to', 'TO'),
('plant', 'NN'),
('50', 'CD')

```

In [63]:

```
#now calculating TF term frequency
```

In [65]:



```
words = [text,text2]
wordset = set()

for i in words:
    temp = i.split()
    for word in temp:
        wordset.add(word)
```

wordset

```
paramati ,
'Besides,',
'Between',
'CSR,',
'Chauhan',
'Circle',
'Colleges,',
'Dara',
'Daund,',
'Defence,',
'Department',
'Department.',
'Division',
'Division,',
'Express',
'For',
'Forest',
'Forest'',
'Green',
'Harit',
...
```

In [75]:



```
import numpy as np
dict = {}

def IDF():
    for it in wordset:
        temp = 0

        for i in range(2):
            if it in words[i].split():
                temp += 1
        dict[it] = np.log10(2/temp)
        print(it,dict[it])
```

IDF()

```
details 0.3010299956639812
water. 0.3010299956639812
vowed 0.3010299956639812
operational 0.3010299956639812
into 0.3010299956639812
comprises 0.3010299956639812
successful 0.3010299956639812
"All 0.3010299956639812
however, 0.3010299956639812
2018 0.3010299956639812
1st 0.3010299956639812
"We 0.3010299956639812
protection 0.3010299956639812
verify 0.3010299956639812
trees 0.0
cent 0.3010299956639812
Society. 0.3010299956639812
rest 0.3010299956639812
report. 0.3010299956639812
it 0.3010299956639812
```

In [81]:



```
document = [text,text2]
dict = {}
for it in document:
    for word in it.split():
        if word in dict:
            dict[word]+=1
        else:
            dict[word] = 1

    for k,v in dict.items():
        dict[k] = v/len(it)

print('Term Frequency')
for k,v in dict.items():
    print(k,v)
```

Term Frequency
Millions 1.1680046159542422e-07
of 0.00580638454683173
people 1.1680046159542422e-07
in 0.002903659475262246
India 3.5040138478627263e-07
took 1.1680046159542422e-07
part 0.00026396904320565874
an 0.0002642026441288496
annual 2.3360092319084845e-07
tree 0.0005281716873345085
planting 0.0007917903291553807
drive 0.0010555257714378488
Sunday. 1.1680046159542422e-07
More 1.1680046159542422e-07
than 1.1680046159542422e-07
250 1.1680046159542422e-07
million 2.3360092319084845e-07
saplings 0.0015840478601571433

In []:

