Lab 2: Text Data Visualization

Objectives:

- To gain more practice in exploring and pre-processing text data.
- To create visualization for the textual data using the techniques introduced in class.

65070501037
Paweekorn Soratyathorn

In [2]:
```
# from google.colab import drive
# drive.mount('/content/drive')
```

Download the data (UN General Debate)

In [1]:
```
# !wget https://github.com/blueprints-for-text-analytics-python/blueprints-text/raw
```

Read the data

In [80]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [81]:
```
df = pd.read_csv('https://github.com/blueprints-for-text-analytics-python/blueprint
                 compression='gzip')
df
```

Out[81]:

	session	year	country	country_name	speaker	position	text
0	25	1970	ALB	Albania	Mr. NAS	NaN	33: May I first convey to our President the co...
1	25	1970	ARG	Argentina	Mr. DE PABLO PARDO	NaN	177.\t : It is a fortunate coincidence that pr...
2	25	1970	AUS	Australia	Mr. McMAHON	NaN	100.\t It is a pleasure for me to extend to y...
3	25	1970	AUT	Austria	Mr. KIRCHSCHLAEGER	NaN	155.\t May I begin by expressing to Ambassado...
4	25	1970	BEL	Belgium	Mr. HARMEL	NaN	176. No doubt each of us, before coming up to ...
...	...	...	...	...	...	...	...
7502	70	2015	YEM	Yemen	Mr. Abdrabuh Mansour Hadi Mansour	President	On behalf of the people and the Government of ...
7503	70	2015	YUG	Yugoslavia	Mr. Tomislav Nikolić	President	\nSeventy years have passed since the establis...
7504	70	2015	ZAF	South Africa	Jacob Zuma	President	I should like to congratulate the President an...
7505	70	2015	ZMB	Zambia	Mr. Edgar Chagwa Lungu	President	I would like to begin by thanking the Secretar...
7506	70	2015	ZWE	Zimbabwe	Robert Mugabe	President	Allow me at the outset to extend to Mr. Mogens...

7507 rows × 7 columns

EDA - Explore more about this dataset

Add new column which presents length of the text

In [5]:
```
df['length'] = df['text'].str.len()
df
```

Out[5]:

	session	year	country	country_name	speaker	position	text	length
0	25	1970	ALB	Albania	Mr. NAS	NaN	33: May I first convey to our President the co...	51419
1	25	1970	ARG	Argentina	Mr. DE PABLO PARDO	NaN	177.\t : It is a fortunate coincidence that pr...	29286
2	25	1970	AUS	Australia	Mr. McMAHON	NaN	100.\t It is a pleasure for me to extend to y...	31839
3	25	1970	AUT	Austria	Mr. KIRCHSCHLAEGER	NaN	155.\t May I begin by expressing to Ambassado...	26616
4	25	1970	BEL	Belgium	Mr. HARMEL	NaN	176. No doubt each of us, before coming up to ...	25911
...	...	...	...	...	...	...	...	...
7502	70	2015	YEM	Yemen	Mr. Abdrabuh Mansour Hadi Mansour	President	On behalf of the people and the Government of ...	10568
7503	70	2015	YUG	Yugoslavia	Mr. Tomislav Nikolić	President	\nSeventy years have passed since the establis...	25430
7504	70	2015	ZAF	South Africa	Jacob Zuma	President	I should like to congratulate the President an...	13662
7505	70	2015	ZMB	Zambia	Mr. Edgar Chagwa Lungu	President	I would like to begin by thanking the Secretar...	14247
7506	70	2015	ZWE	Zimbabwe	Robert Mugabe	President	Allow me at the outset to extend to Mr. Mogens...	11013

7507 rows × 8 columns

List all columns in this dataset

In [6]:
```
df.columns
```

Out[6]:
```
Index(['session', 'year', 'country', 'country_name', 'speaker', 'position',
       'text', 'length'],
      dtype='object')
```

Types of data for each column

In [7]:
```
df.dtypes
```

Out[7]:
```
session          int64
year             int64
country         object
country_name    object
speaker         object
position        object
text            object
length           int64
dtype: object
```

All information

In [8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7507 entries, 0 to 7506
Data columns (total 8 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   session       7507 non-null   int64 
 1   year          7507 non-null   int64 
 2   country       7507 non-null   object
 3   country_name  7507 non-null   object
 4   speaker       7480 non-null   object
 5   position      4502 non-null   object
 6   text          7507 non-null   object
 7   length        7507 non-null   int64 
dtypes: int64(3), object(5)
memory usage: 469.3+ KB
```

For all integer-typed columns, find out their stats

In [9]:
```
df.describe().T
```

Out[9]:

	count	mean	std	min	25%	50%	75%	max
session	7507.0	49.610763	12.892155	25.0	39.0	51.0	61.0	70.0
year	7507.0	1994.610763	12.892155	1970.0	1984.0	1996.0	2006.0	2015.0
length	7507.0	17967.281604	7860.038463	2362.0	12077.0	16424.0	22479.5	72041.0

Investigate: Any missing data?

In [10]:
```
# Missing data
df.isna().sum()
```

Out[10]:
```
session            0
year               0
country            0
country_name       0
speaker           27
position        3005
text               0
length             0
dtype: int64
```

Address the missing data in the column "speaker" by replacing those missing values with 'unknown'.

In [11]:
```
df['speaker'].fillna('unknown',inplace=True)
df.isna().sum()
```

Out[11]:
```
session            0
year               0
country            0
country_name       0
speaker            0
position        3005
text               0
length             0
dtype: int64
```

Find out all unique speakers in this dataset.

In [12]:
```
pd.unique(df['speaker'])
```

Out[12]:
```
array(['Mr. NAS', 'Mr. DE PABLO PARDO', 'Mr. McMAHON', ...,
       'Mr. Abdrabuh Mansour Hadi Mansour', 'Mr. Tomislav Nikolić',
       'Mr. Edgar Chagwa Lungu'], dtype=object)
```

In [13]:
```
len(pd.unique(df['speaker']))
```

Out[13]:
```
5429
```

Filter only records of which the speaker is President 'Bush'.

In [14]:
```
df[df['speaker'].str.contains('Bush')]
```

Out[14]:

	session	year	country	country_name	speaker	position	text	length
2720	44	1989	USA	United States	Bush	President	I am honoured to address the General Assembly...	19779
3038	46	1991	USA	United States	George Bush	President	I am honoured to speak with you as you open t...	15555
4814	56	2001	USA	United States	George W. Bush	President	We meet in a Hall devoted to\npeace; in a cit...	14724
5002	57	2002	USA	United States	Mr. George W. Bush	President	We meet one year and one day\nafter a terrori...	16684
5191	58	2003	USA	United States	George W. Bush	President	Twenty-four months ago, and\nyesterday in the...	16962
5382	59	2004	USA	United States	Mr. George W Bush	President	Thank you for the honour of\naddressing the Ge...	18628
5760	61	2006	USA	United States	Mr. George W. Bush	President	I am grateful for the privilege of \nspeaking ...	16675
5951	62	2007	USA	United States	George W. Bush	President	Thank you for the opportunity \nto address the...	15483
6143	63	2008	USA	United States	George W. Bush	President	I am pleased to be here to \naddress the Gener...	18384

Select text from a particular row

In [15]:
```
temp = df[df['speaker'].str.contains('Bush')]
temp.iloc[[2]].text.values
```

Out[15]:
array(['\ufeffWe meet in a Hall devoted to\npeace; in a city scarred by violence; in a nation\nawakened to danger; in a world uniting for a long\nstruggle. Every civilized nation here today is resolved\nto keep the most basic commitment of civilization. We\nwill defend ourselves and our future against terror and\nlawless violence.\nThe United Nations was founded in this cause. In\nthe Second World War, we learned that there is no\nisolation from evil. We affirmed that some crimes are\nso terrible they offend humanity itself, and we resolved\nthat the aggressions and ambitions of the wicked must\nbe opposed early, decisively and collectively, before \nthey threaten us all.\nThat evil has returned, and that cause is renewed.\nA few miles from here, many thousands still lie in a\ntomb of rubble. Tomorrow, the Secretary-General, the\nPresident of the General Assembly and I will visit that\nsite, where the names of every nation and region that\nlost citizens will be read aloud. If we were to read out\nthe names of every person who died, it would take\nmore than three hours.\nThose names include a citizen of the Gambia,\nwhose wife spent their fourth wedding anniversary, 12\nSeptember, searching in vain for her husband. Those\nnames include a man who supported his wife in\nMexico, sending home money every week. Those\nnames include a young Pakistani who prayed towards\nMecca five times a day and who died that day trying to\nsave others.\nThe suffering of 11 September was inflicted on\npeople of many faiths and many nations. All of the\nvictims, including Muslims, were killed with equal\nindifference and equal satisfaction by the terrorist\nleaders.\nThe terrorists are violating the tenets of every\nreligion, including the one they invoke. Last week, the\nsheikh of Al-Azhar University, the world’s oldest\nIslamic institution of higher learning, declared that\nterrorism is a disease and that Islam prohibits killing\ninnocent civilians. The terrorists call their cause holy,\nyet they fund it with drug dealing. They encourage\nmurder and suicide in the name of a great faith that\nforbids both. They dare to ask God’s blessing as they\nset out to kill innocent men, women and children. But\n8\n\nthe God of Isaac and Ishmael would never answer such\na prayer. And a murderer is not a martyr; he is just a\nmurderer.\nTime is passing. Yet for the United States of\nAmerica, there will be no forgetting 11 September. We\nwill remember every rescuer who died in honour. We\nwill remember every family that lives in grief. We will\nremember the fire and ash, the last phone calls, the\nfunerals of the children.\nAnd the people of my country will remember\nthose who have plotted against us. We are learning\ntheir names. We are coming to know their faces. There\nis no corner of the Earth distant or dark enough to\nprotect them. However long it takes, their hour of\njustice will come.\nEvery nation has a stake in this cause. As we\nmeet, the terrorists are planning more murder —\nperhaps in my country, or perhaps, fellow members, in\nyours. They kill because they aspire to dominate. They\nseek to overthrow Governments and to destabilize\nentire regions. Last week, anticipating this meeting of\nthe General Assembly, they denounced the United\nNations; they called our Secretary-General a criminal\nand they condemned all Arab nations here as traitors to\nIslam. Few countries meet their exacting standards of\nbrutality and oppression. Every other country is a\npotential target.\nAnd all the world faces the most horrifying\nprospect of all: those same terrorists are searching for\nweapons of mass destruction, the tools to turn their\nhatred into holocaust. They can be expected to use\nchemical, biological and nuclear weapons the moment\nthey are capable of doing so. No hint of conscience\nwould prevent it. That threat cannot be ignored; that\nthreat cannot be appeased. Civilization itself — the\ncivilization we share — is threatened. History will\nrecord our response and will judge or justify every\nnation in this Hall.\nThe civilized world is now responding. We act to\ndefend ourselves and to deliver our children from a\nfuture of fear. We choose the dignity of life over a\nculture of death. We choose lawful change and civil\ndisagreement over coercion, subversion and chaos.\nThose commitments — hope and order, law and life —\nunite people across cultures and continents. Upon those \ncommitments depend all peace and progress. For those\ncommitments, we are determined to fight.\nThe United Nations has risen to this\nresponsibility: on 12 September, these buildings\nopened for emergency meetings of the General\nAssembly and of the Security Council. Before the sun\nhad set, these attacks on the world stood condemned by\nthe world, and I want to thank you, fellow members,\nfor that strong and principled stand.\nI also thank the Arab and Islamic countries that\nhave condemned terrorist murder. Many of you have\nseen the destruction of terror in your own lands. The\nterrorists are increasingly isolated by their own hatred\nand extremism. They cannot hide behind Islam. The\nauthors of mass murder and their allies have no place\nin any culture, and no home in any faith.\nThe conspiracies of terr

or are being answered by\nan expanding global coalition. Not every nation will be \npart of every action against the enemy. But every\nnation in our coalition has d uties. Those duties can be\ndemanding, as we in America are learning. We have\nalr eady made adjustments in our laws and in our daily\nlives. We are taking new measu res to investigate terror\nand to protect against threats.\nThe leaders of all nat ions must now carefully\nconsider their responsibilities and their future. Terrori st\ngroups such as Al Qaeda depend upon the aid or\nindifference of Governments. T hey need the support of\na financial infrastructure and safe havens to train and\n plan and hide.\nSome nations want to play their part in the fight\nagainst terror but tell us they lack the means to enforce\ntheir laws and control their borders. We stand ready to\nhelp.\nSome Governments still turn a blind eye to the\nterroris ts, hoping the threat will pass them by. They are\nmistaken.\nAnd some Government s, while pledging to\nuphold the principles of the United Nations, have cast\nthei r lot with the terrorists. They support them and\nharbour them. And they will find that their welcomed\nguests are parasites that will weaken them and\neventually co nsume them. For every regime that\nsponsors terror there is a price to be paid, an d it will be\npaid. The allies of terror are equally guilty of murder\nand equally accountable to justice.\nThe Taliban are now learning that lesson. That\nregime an d the terrorists who support it are now\nvirtually indistinguishable. Together the y promote\n9\n\nterror abroad and impose a reign of terror on the\nAfghan people. Women are executed in Kabul’s soccer\nstadium. They can be beaten for wearing sock s that are\ntoo thin. Men are jailed for missing prayer meetings.\nThe United Stat es, supported by many nations, is\nbringing justice to the terrorists in Afghanist an. We are\nmaking progress against military targets – and that is\nour objective. Unlike the enemy, we seek to minimize,\nnot maximize, the loss of innocent life. I am proud of\nthe honourable conduct of the American military. And\nmy country grie ves for all the suffering the Taliban\nhave brought upon Afghanistan, including th e terrible\nburden of war.\nThe Afghan people do not deserve their present\nruler s. Years of Taliban misrule have brought nothing\nbut misery and starvation. Even before this current\ncrisis, 4 million Afghans depended on food from the\nUnited S tates and other nations, and millions of\nAfghans were refugees from Taliban oppre ssion.\nI make this promise to all the victims of that\nregime: the Taliban’s days of harbouring terrorists, and\ndealing in heroin, and brutalizing women are drawin g\nto a close. When that regime is gone, the people of\nAfghanistan will say, with the rest of the world, good\nriddance.\nI can promise, too, that America will join the\nworld in helping the people of Afghanistan rebuild\ntheir country.\nMany nati ons, including mine, are sending food\nand medicine to help Afghans through the wi nter.\nAmerica has airdropped over 1.3 million packages of\nrations in Afghanista n. Just this week, we airlifted\n20,000 blankets and over 200 tons of provisions i nto\nthe region. We continue to provide humanitarian aid,\neven while the Taliban try to steal the food we send.\nMore help, eventually, will be needed. The\nUnited States will work closely with the United\nNations and development banks to reconst ruct\nAfghanistan after hostilities there have ceased and the\nTaliban are no long er in control. And the United States\nwill work with the United Nations to support a post-\nTaliban Government that represents all of the Afghan\npeople.\nIn this wa r of terror, each of us must answer for\nwhat we have done or what we have left un done. After\ntragedy, there is a time for sympathy and condolence.\nMy country has been very grateful for both.\nThe memorials and vigils around the world will\nnot be forgotten. But the time for sympathy has now\npassed. The time for action has n ow arrived.\nThe most basic obligations in this new conflict\nhave already been de fined by the United Nations. On\n28 September, the Security Council adopted resolu tion\n1373 (2001). Its requirements are clear: every\nUnited Nations Member has a responsibility to crack\ndown on terrorist financing. We must pass all necessary\n laws in our own countries to allow the confiscation of\nterrorist assets. We must apply those laws to every\nfinancial institution in every nation.\nWe have a respo nsibility to share intelligence and\ncoordinate the efforts of law enforcement. If you know\nsomething, tell us. If we know something, we will tell\nyou. And when we find the terrorists, we must work\ntogether to bring them to justice.\nWe have a r esponsibility to deny any sanctuary,\nsafe haven, or transit to terrorists. Every known\nterrorist camp must be shut down, its operators\napprehended, and evidence of their arrest presented to\nthe United Nations.\nWe have a responsibility to den y weapons to\nterrorists – and to actively prevent private citizens\nfrom providin g them.\nThese obligations are urgent, and they are binding\non every nation with a place in this Hall. Many\nGovernments are taking these obligations seriously,\na

```
nd my country appreciates it. Yet even beyond\nresolution 1373 (2001), more is req
uired – and more\nis expected – of our coalition against terror. We are\nasking fo
r a comprehensive commitment to this fight.\nWe must unite in opposing all terrori
sts, not just\nsome of them. In this world, there are good causes and\nbad causes,
and we may disagree on where that line is\ndrawn. Yet there is no such thing as a
good terrorist.\nNo national aspiration, no remembered wrong, can ever\njustify th
e deliberate murder of the innocent. Any\nGovernment that rejects this principle –
trying to pick\nand choose its terrorist friends – will know the\nconsequences.\nW
e must speak the truth about terror. Let us never\ntolerate outrageous conspiracy
theories concerning the\nattacks of 11 September – malicious lies that attempt\nto
shift the blame away from the terrorists themselves,\naway from the guilty. To inf
lame ethnic hatred is to\nadvance the cause of terror.\n10\n\nThe war against terr
or must not serve as an\nexcuse to persecute ethnic and religious minorities in\na
ny country. Innocent people must be allowed to live\ntheir own lives, by their own
customs, under their own\nreligion. And every nation must have avenues for the\npe
aceful expression of opinion and dissent. When these\navenues are closed, the temp
tation to speak through\nviolence grows.\nWe must press on with our agenda for pea
ce and\nprosperity in every land. My country is pledged to\nencouraging developmen
t and expanding trade. My\ncountry is pledged to investing in education and\ncomba
ting AIDS and other infectious diseases around\nthe world. Following 11 September,
these pledges are\neven more important. In our struggle against hateful\ngroups th
at exploit poverty and despair, we must offer\nan alternative of opportunity and h
ope.\nThe American Government also stands by its\ncommitment to a just peace in th
e Middle East. We are\nworking towards a day when two States – Israel and\nPalesti
ne – live peacefully together, within secure and\nrecognized borders, as called fo
r by the Security\nCouncil resolutions. We will do all in our power to\nbring both
parties back into negotiations. But peace\nwill come only when all have sworn off
– forever –\nincitement, violence and terror.\nFinally, this struggle is a definin
g moment for the\nUnited Nations itself – and the world needs its\nprincipled lead
ership. It undermines the credibility of\nthis great institution, for example, whe
n the\nCommission on Human Rights offers seats to some of\nthe world's most persis
tent violators of human rights.\nThe United Nations depends, above all, on its mor
al\nauthority – and that authority must be preserved.\nThe steps I have described
will not be easy. For\nall nations, they will require effort. For some nations,\nt
hey will require great courage. Yet the cost of inaction\nis far greater. The only
alternative to victory is a\nnightmare world where every city is a potential killi
ng\nfield.\nAs I told the American people, freedom and fear\nare at war. We face e
nemies that hate not our policies,\nbut our existence – the tolerance of openness
and the\ncreative culture that define us. But the outcome of this\nconflict is cer
tain.\nThere is a current in history, and it runs towards\nfreedom. Our enemies re
sent it and dismiss it, but the\ndreams of mankind are defined by liberty – the na
tural\nright to create, and build, and worship, and live in\ndignity. When men and
women are released from\noppression and isolation, they find fulfilment and hop
e,\nand they leave poverty by the millions. These\naspirations are lifting up the
peoples of Europe, Asia,\nAfrica and the Americas, and they can lift up all of the
\nIslamic world. We stand for the permanent hopes of\nhumanity, and those hopes wi
ll not be denied.\nWe are confident, too, that history has an author,\nwho fills t
ime and eternity with his purpose. We know\nthat evil is real, but good will preva
il against it. This is\nthe teaching of many faiths. And in that assurance, we\nga
in strength for a long journey.\nIt is our task – the task of this generation – to
\nprovide the response to aggression and terror. We have\nno other choice, because
there is no other peace. We\ndid not ask for this mission, yet there is honour in
\nhistory's call. We have a chance to write the story of\nour times – a story of c
ourage defeating cruelty, and\nlight overcoming darkness. This calling is worthy o
f\nany life, and worthy of every nation. So let us go\nforward – confident, determ
ined and unafraid.\n\n\n\n'],
      dtype=object)
```

Tasks

Select 3 speeches from different people and perform the following:

- Pre-process each speech using techniques taught in class
- Find top 10 words
- Select proper text visualization technique (at least 2 techniques) to gain more insights about the speech
- Briefly describe the insights you gain from the visualizations that you create

Preprocessing

1. tokenizer

In [16]:
```
from nltk.tokenize import word_tokenize

people = ['Teodoro Obiang Nguema Mbasogo', 'Kenny D. Anthony', 'Prince Norodom Siri
filter_df = df[ df['speaker'].isin(people) ]
speeches_df = filter_df.copy().groupby('speaker').head(1)  # filter only 1 speech f

speeches_df['tokens'] = [word_tokenize(text) for text in speeches_df['text']]
speeches_df
```

Out[16]:

	session	year	country	country_name	speaker	position	text	length	toke
3479	49	1994	KHM	Cambodia	Prince Norodom Sirivudh	Deputy Prime Minister	On this solemn occasion allow me first to\ncon...	12705	[On, th solem occasi allow, n firs
4015	52	1997	LCA	Saint Lucia	Kenny D. Anthony	Prime Minister	My delegation welcomes\nthe experience and ex...	19785	[M delegati welcom th experien
4343	54	1999	GNQ	Equatorial Guinea	Teodoro Obiang Nguema Mbasogo	President	At the outset, we would like to congratulate y...	15087	[At, th outset we, wou like, cong

1. stop-words removal

In [54]:
```
from nltk.corpus import stopwords

speeches = list(speeches_df['tokens'])
stop_words = stopwords.words('english')
filtered_speech = [ [], [], [] ]

for ind, word_list in enumerate(speeches):
    filtered_speech[ind] = [words for words in word_list if words.lower() not in st
    print(filtered_speech[ind].__contains__(stop_words))
```

```
False
False
False
```

1. lemmatization/ stemming

In [59]:
```
from nltk.stem import WordNetLemmatizer

lemmatizer = WordNetLemmatizer()
speech_1 = filtered_speech[0]
speech_2 = filtered_speech[1]
speech_3 = filtered_speech[2]

for speech in [speech_1, speech_2, speech_3]:
    for word in speech:
        temp = []
        temp.append(lemmatizer.lemmatize(word))

    speech = temp
```

1. Special Characters/ Punctuation

In [60]:
```
import re

def filter_alphanumeric(speech):
    full_speech = ' '.join(speech)
    text_prep = re.sub(r'[^a-zA-Z0-9\s]', '', full_speech)
    return text_prep


cleaned_speech_1 = filter_alphanumeric(speech_1)
cleaned_speech_2 = filter_alphanumeric(speech_2)
cleaned_speech_3 = filter_alphanumeric(speech_3)
```

Text Analysis

1. Prince Norodom Sirivudh

- Find top 10 words

In [78]:
```
def calculate_word_frequencies(words):
    # Calculate word frequencies
    word_freq = {}
    for word in words:
        word_freq[word] = word_freq.get(word, 0) + 1
    return word_freq

# Tokenize the text
def top10_words(speech):
    words = word_tokenize(speech)
    filtered_sentence = []

    for w in words:
        if w not in stop_words:
            filtered_sentence.append(w)

    # Calculate word frequencies
    word_freq = calculate_word_frequencies(filtered_sentence)

    # Convert word frequencies to a DataFrame for seaborn
    data = {'Word': list(word_freq.keys()), 'Frequency': list(word_freq.values())}
    df_word_freq = pd.DataFrame(data)

    # Sort DataFrame by frequency in descending order
    df_word_freq = df_word_freq.sort_values(by='Frequency', ascending=False)
```

```
    return df_word_freq.head(10)


top10_words(cleaned_speech_1)
```

Out[78]:

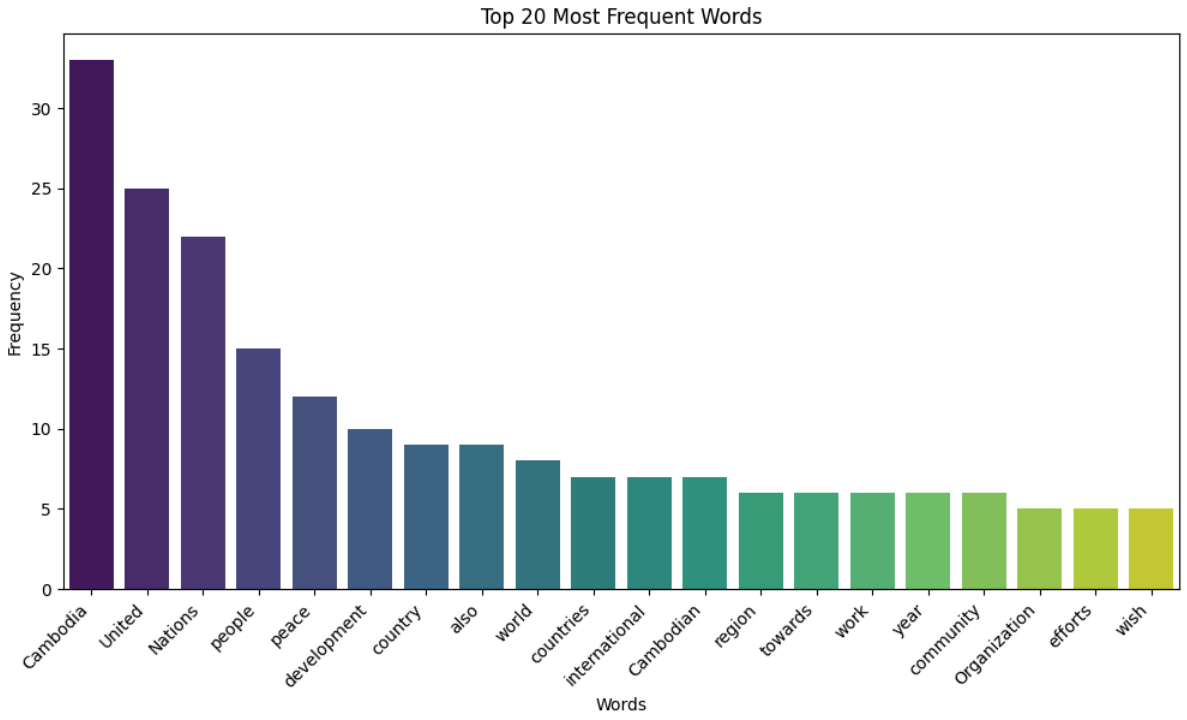
	Word	Frequency
66	Cambodia	33
13	United	25
14	Nations	22
102	people	15
35	peace	12
131	development	10
34	country	9
27	also	9
61	world	8
145	countries	7

- Text Visualization

In [83]:

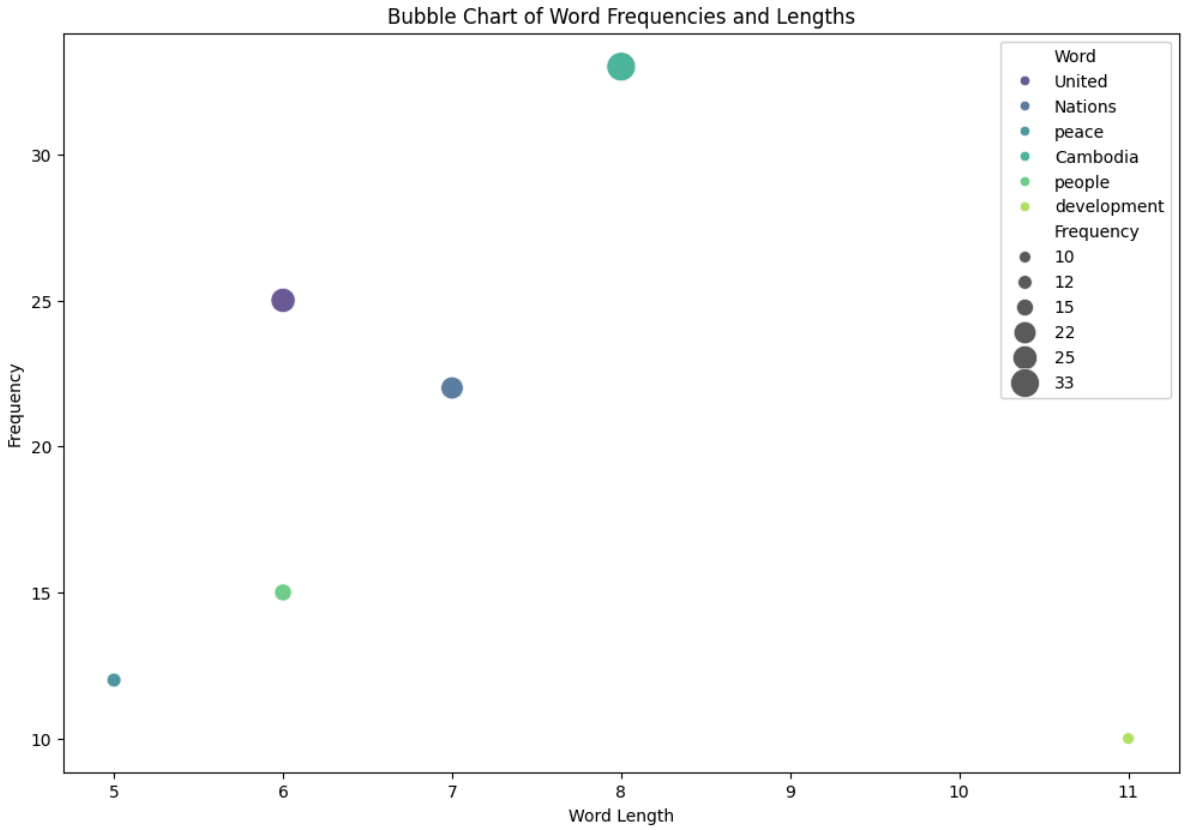
```
from wordcloud import WordCloud

def wordcloud_plot(speech):
    # Generate word cloud
    wordcloud = WordCloud(width=800, height=400, background_color='white').generate

    # Display the generated word cloud using matplotlib
    plt.figure(figsize=(10, 5))
    plt.imshow(wordcloud, interpolation='bilinear')
    plt.axis('off')
    plt.show()


wordcloud_plot(cleaned_speech_1)
```

In [70]:
```
from nltk.probability import FreqDist

def frequency_plot(speech):
    # tokenize
    words = word_tokenize(speech)

    # Calculate word frequencies
    word_freq = FreqDist(words)

    # Convert word frequencies to a DataFrame for seaborn
    data = {'Word': list(word_freq.keys()), 'Frequency': list(word_freq.values())}
    df_word_freq = pd.DataFrame(data)

    # Sort DataFrame by frequency in descending order
    df_word_freq = df_word_freq.sort_values(by='Frequency', ascending=False)

    # Plot a bar chart using seaborn
    plt.figure(figsize=(12, 6))
    sns.barplot(x='Word', y='Frequency', data=df_word_freq.head(20), palette='viric
    plt.title('Top 20 Most Frequent Words')
    plt.xlabel('Words')
    plt.ylabel('Frequency')
    plt.xticks(rotation=45, ha='right')  # Rotate x-axis labels for better readabil
    plt.show()


frequency_plot(cleaned_speech_1)
```



In [76]:
```
# Tokenize the text into words
words = word_tokenize(cleaned_speech_1)

# Calculate word frequencies
word_freq = FreqDist(words)

# Create a DataFrame with word frequencies and lengths
data = {'Word': list(word_freq.keys()), 'Frequency': list(word_freq.values()), 'Ler
df_word_data = pd.DataFrame(data)

# Filter out words with frequency less than 2 for better visualization
df_word_data = df_word_data[df_word_data['Frequency'] >= 10]
```

```
# Plot a bubble chart using seaborn
plt.figure(figsize=(12, 8))
sns.scatterplot(x='Length', y='Frequency', size='Frequency', data=df_word_data, hue
plt.title('Bubble Chart of Word Frequencies and Lengths')
plt.xlabel('Word Length')
plt.ylabel('Frequency')
plt.show()
```



Insight

Prince Norodom Sirivudh is related with Cambodia and we can imply from charts and word frequency that he is the person who has responsible with peace and development of Cambodia. After I researched about his profile, I found that he is a Cambodian politician.

2. Kenny D. Anthony

- Find top 10

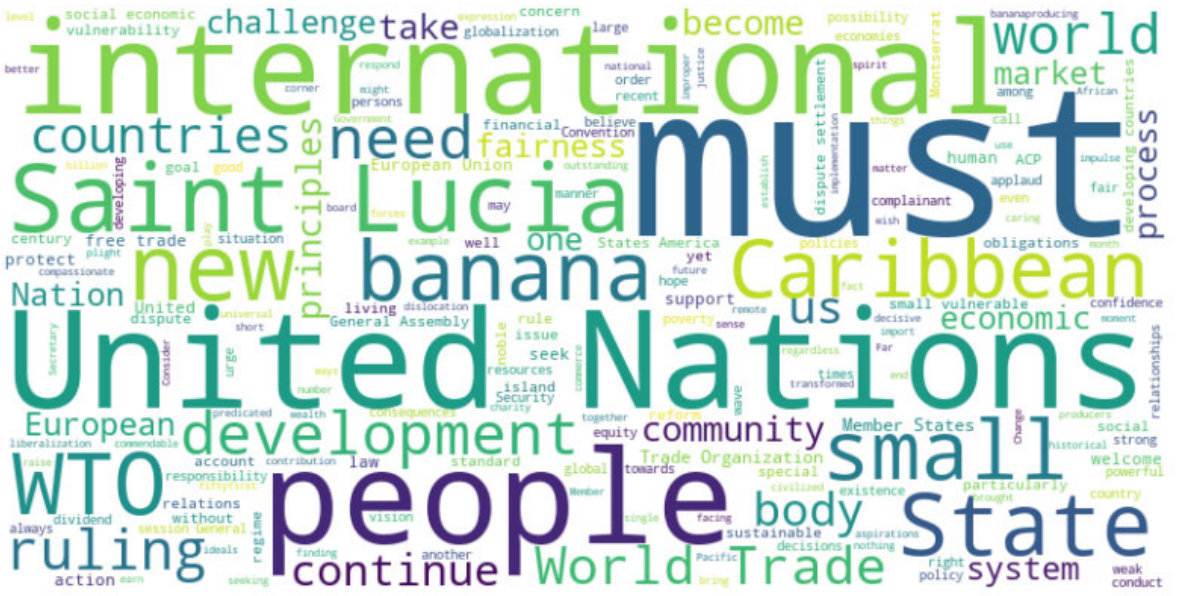
In [79]:
```
top10_words(cleaned_speech_2)
```

Out[79]:

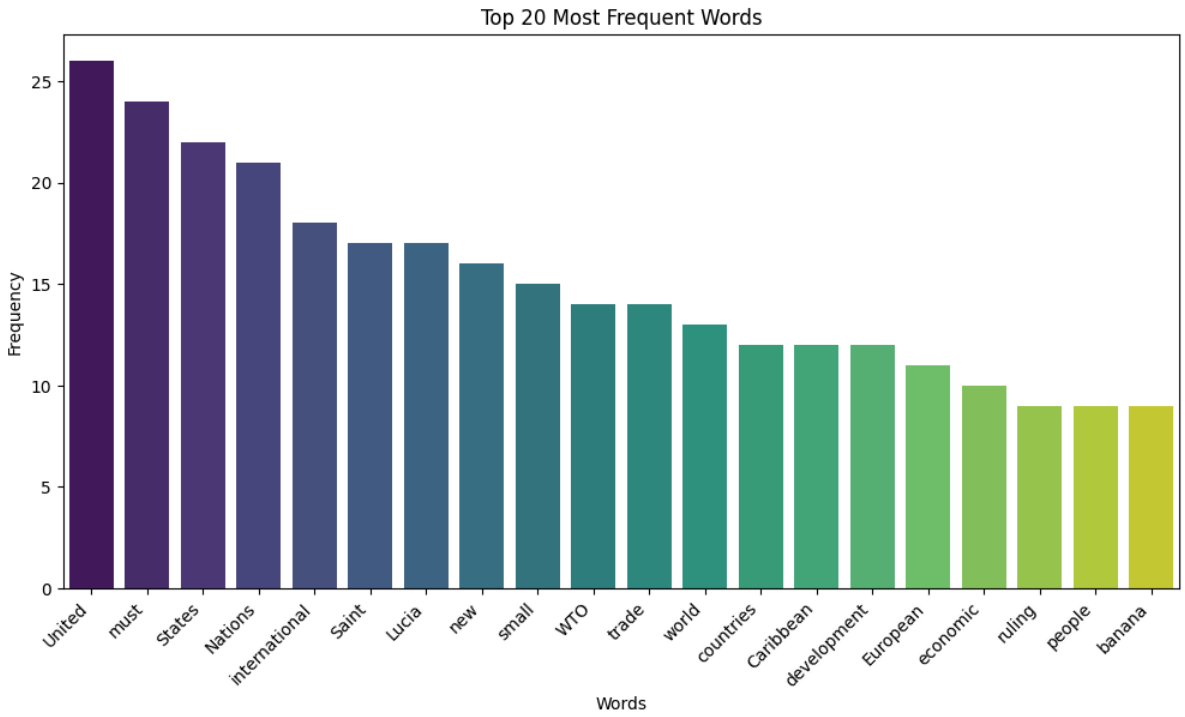
	Word	Frequency
33	United	26
133	must	24
177	States	22
34	Nations	21
225	international	18
157	Saint	17
158	Lucia	17
31	new	16
61	small	15
286	WTO	14

- Text Visualization

In [84]:
```
wordcloud_plot(cleaned_speech_2)
```



In [71]:
```
frequency_plot(cleaned_speech_2)
```



Insight

I implied that Mr. Kenny D. Anthony is an influence in economics development and international trade of United States from the word *WTO*. I am not quite sure what is the main product he trade between countries so, I implied from the bar charts that it is banana. After researched for his profile, he is a Saint Lucian politician who owns banana industry.

3. Teodoro Obiang Nguema Mbasogo

- Find Top 10

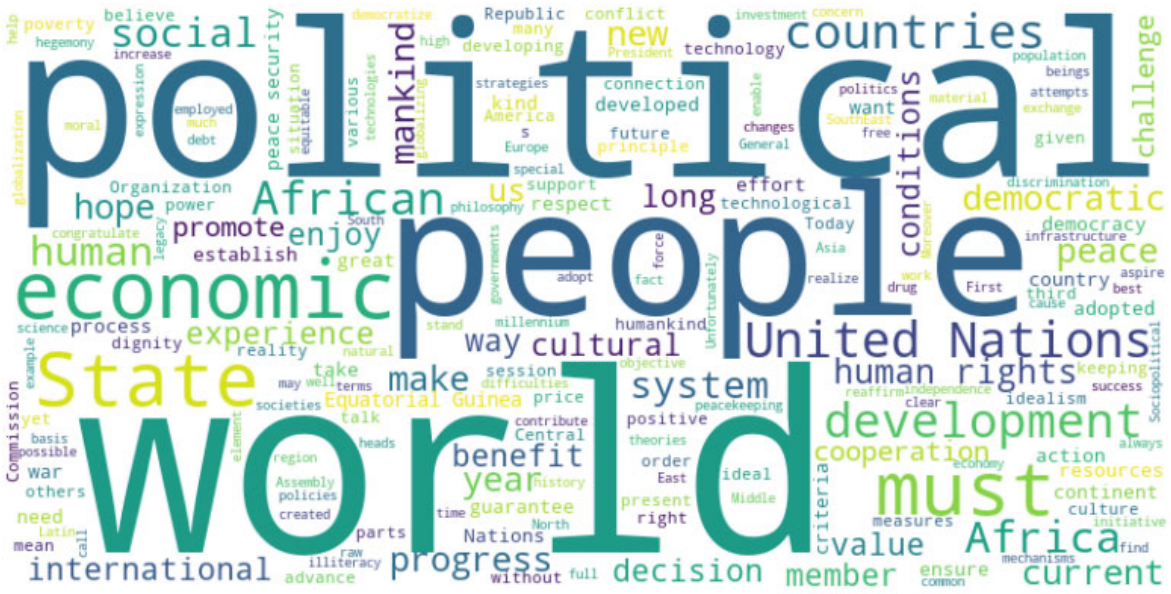
In [82]:
```
top10_words(cleaned_speech_3)
```

Out[82]:

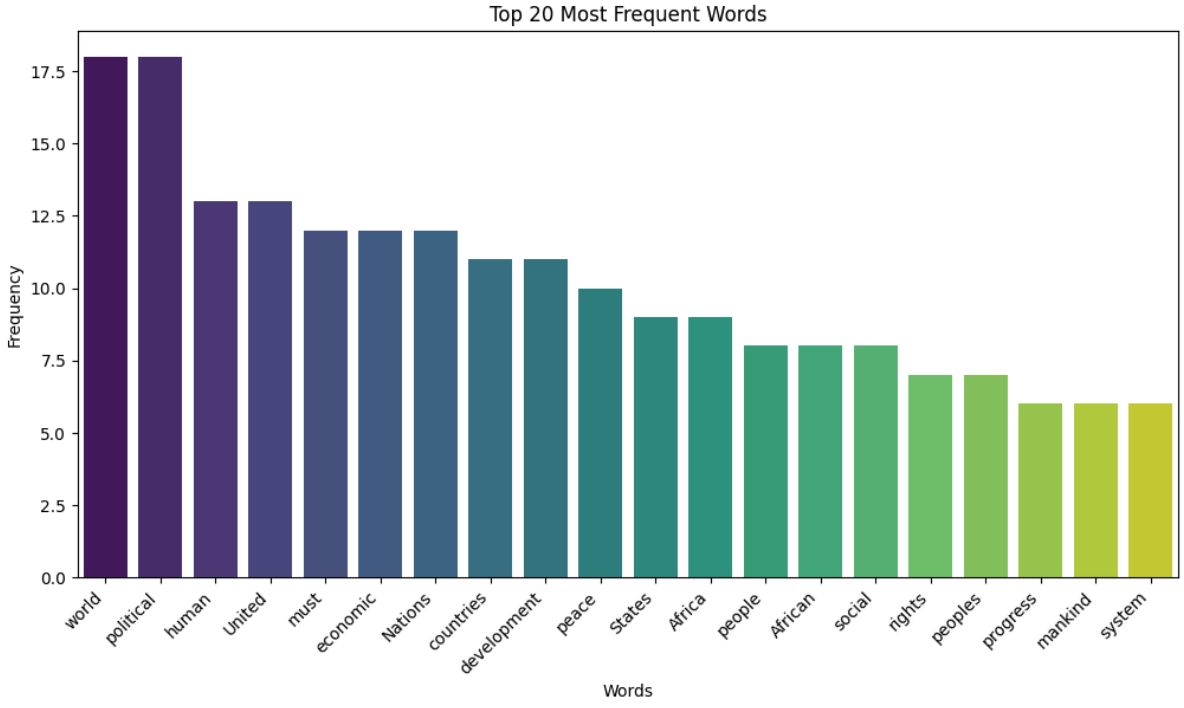
	Word	Frequency
67	world	18
80	political	18
121	human	13
12	United	13
154	must	12
81	economic	12
13	Nations	12
160	countries	11
178	development	11
104	peace	10

- Text Visualization

In [85]:
```
wordcloud_plot(cleaned_speech_3)
```



In [86]:
```
frequency_plot(cleaned_speech_3)
```



Insight

From the visualization, we can guess that Teodoro Obiang Nguema Mbasogo is an African politician who deals with the economic and peace but, I am not sure about *mankind* and *system* word. The insight I get from the visualization is close to Prince Norodom from my first chosen person. The information from wikipedia shows that he is Equatoguinean politician and dictator.

In [ ]:
```
```