

**Question 1**

1. It imports necessary libraries: **nltk** for natural language processing, **stopwords** from **nltk.corpus** to remove common words, **WordCloud** from the **wordcloud** library to generate a word cloud, and **matplotlib.pyplot** for visualization.
2. The text of the Universal Declaration of Human Rights is read from a file called "un\_declaration\_hr\_text\_data.txt" using the **open()** function and stored in the **declaration\_text** variable.
3. The text is tokenized into individual words using the **word\_tokenize()** function from **nltk**.
4. Stop words, which are common words like "the" and "is" that do not carry significant meaning, are loaded using **stopwords.words("english")** from **nltk.corpus**. These stop words are filtered out from the tokens.
5. A frequency distribution is created using **FreqDist** from **nltk** to count the occurrences of each word in the filtered tokens.
6. A word cloud is generated from the frequency distribution using **WordCloud** with customized settings for width, height, and background color.
7. The word cloud is displayed using **imshow()** from **matplotlib.pyplot**.
8. The top 25 frequent terms from the frequency distribution are extracted using **most\_common()**.
9. The labels (terms) and values (frequencies) are separated into two lists for plotting.
10. A bar plot is created using **plt.bar()** from **matplotlib.pyplot**, displaying the top 25 frequent terms on the x-axis and their corresponding frequencies on the y-axis.
11. The resulting word cloud and bar plot are shown using **plt.show()**.

Overall, the code reads the text, preprocesses it by removing stop words, calculates the frequency distribution of the remaining words, generates a word cloud visualization, and plots a bar chart of the top 25 frequent terms.

import nltk

from nltk.corpus import stopwords

from wordcloud import WordCloud

import matplotlib.pyplot as plt

# Read the text from the file

with open("un\_declaration\_hr\_text\_data.txt", "r") as file:

declaration\_text = file.read()

# Tokenize the text into words

tokens = nltk.word\_tokenize(declaration\_text)

# Filter out stopwords

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

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

# Create a frequency distribution of the tokens

freq\_dist = nltk.FreqDist(filtered\_tokens)

# Generate a word cloud

wordcloud = WordCloud(width=800, height=400, background\_color="white").generate\_from\_frequencies(freq\_dist)

# Plot the word cloud

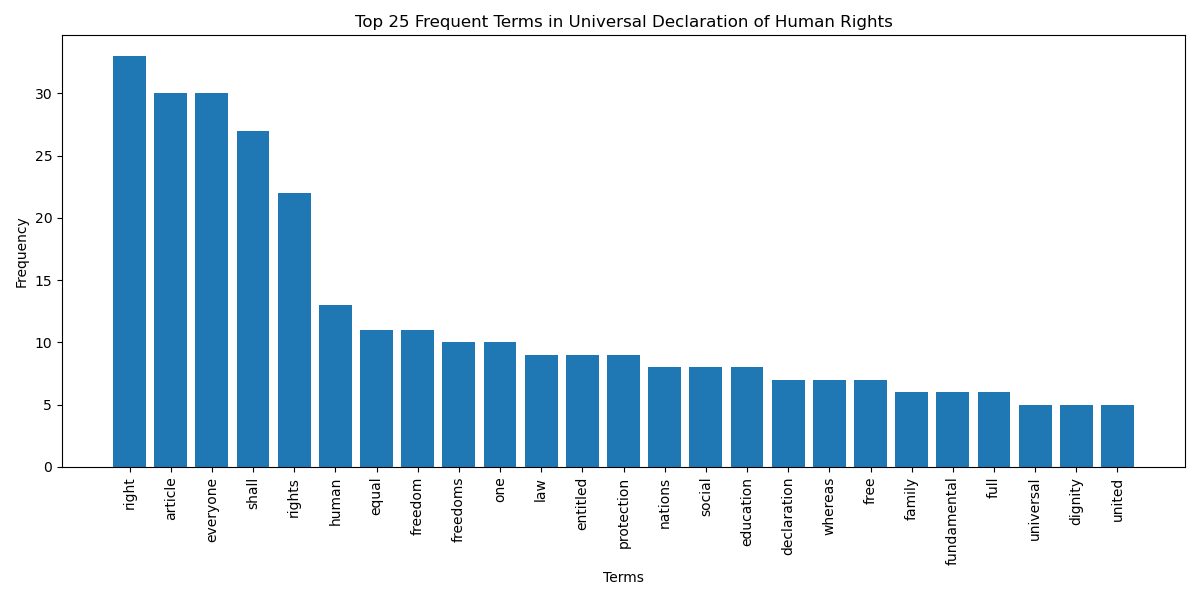
plt.figure(figsize=(10, 5))

plt.imshow(wordcloud, interpolation="bilinear")

plt.axis("off")

plt.title("Word Cloud of Universal Declaration of Human Rights")

plt.show()



# Get the top 25 frequent terms

top\_25 = freq\_dist.most\_common(25)

# Prepare data for bar plot

labels, values = zip(\*top\_25)

# Plot the bar chart

plt.figure(figsize=(12, 6))

plt.bar(labels, values)

plt.xticks(rotation=90)

plt.xlabel("Terms")

plt.ylabel("Frequency")

plt.title("Top 25 Frequent Terms in Universal Declaration of Human Rights")

plt.tight\_layout()

plt.show()

A close-up of words

Description automatically generated with low confidence

Sources: (Jonathan P. Scaccia, 2021) (Diksha Khurana, 2022), (Gruetzemacher, 2022), (Nadkarni, 2011), (Cohen, 2014), (Ashton Pike, 2023), (Le Glaz A, 2021), (Sethunya R Joseph, 2016), (Ronan Collobert, 2000), (van Erp M, 2021), (Chowdhury, 2003), (Santosh K Behera, 2020), (Mah & Skalna, 2022)