Eduvos
Your Education. Your Future.

Individual Assessment Coversheet

To be attached to the front of the assessment.

Campus:	Midrand
Faculty:	Information Technology
Group:	1
Lecturer's Name:	Anyway Chare
Student Full Name:	Keabetsoe Zulu
Student Number:	5WJDJ2YH6

Indicate	Yes	No
Plagiarism report attached		

Declaration:

I declare that this assessment is my own original work except for source material explicitly acknowledged. I also declare that this assessment or any other of my original work related to it has not been previously, or is not being simultaneously, submitted for this or any other course. I also acknowledge that I am aware of the Institution's policy and regulations on honesty in academic work as set out in the Eduvos Conditions of Enrolment, and of the disciplinary guidelines applicable to breaches of such policy and regulations.

Signature	Date

Lecturer's Comments:

Marks Awarded:	%

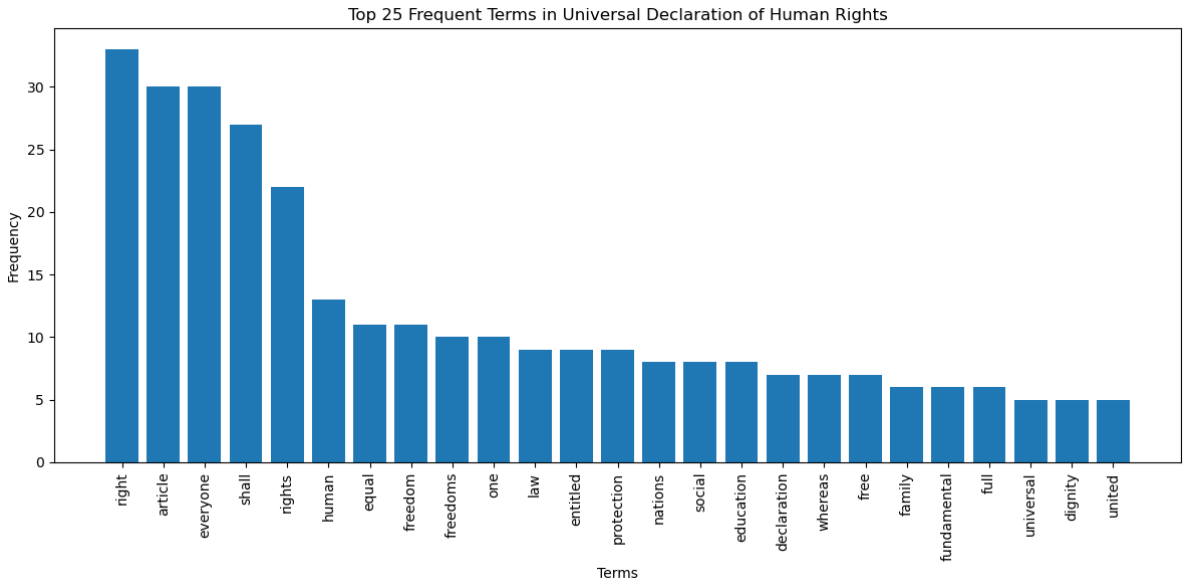
Signature	Date

Question 1

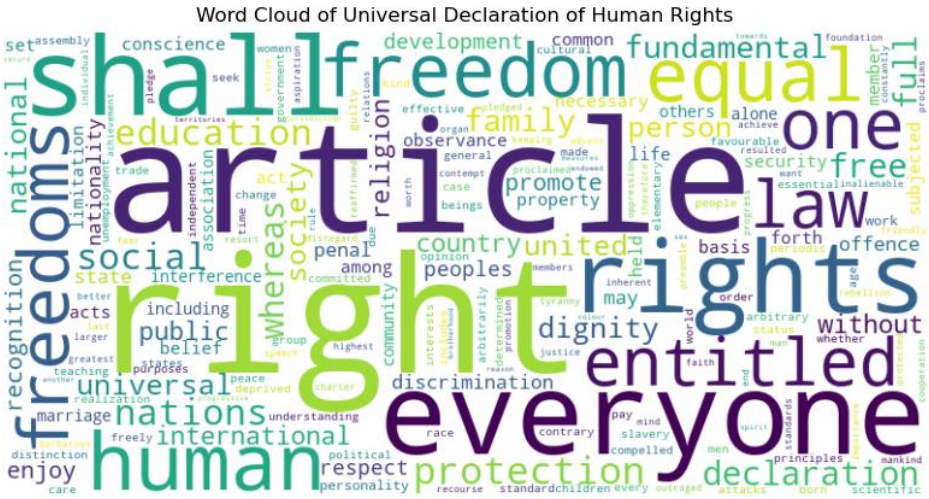
1. It imports necessary libraries: **nltk** for natural language processing, **stopwords** from **nltk.corpus** to remove common words, **WordCloud** from the **wordcloud** library to generate a word cloud, and **matplotlib.pyplot** for visualization.
2. The text of the Universal Declaration of Human 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()
```



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)

References

Ashton Pike, B. B. S. O. G. S. H., 2023. Chapter 9 - Artificial intelligence and machine learning. *Substance Use and Addiction Research,* pp. 99-106.

Chowdhury, G., 2003. Natural language processing. *Annual Review of Information Science and Technology,* Volume 37, pp. 51-89.

Cohen, K. B., 2014. Chapter 6 - Biomedical Natural Language Processing and Text Mining. *Methods in Biomedical Informatics,* pp. 141-177.

Diksha Khurana, A. K. K. K. a. S. S., 2022. Natural Language Processing: State of The Art, Current Trends and Challenges. *Multimedia Tools and Applications,* 82(6), pp. 3713-3744.

Gruetzemacher, R., 2022. *The Power of Natural Language Processing.* [Online]
Available at: https://hbr.org/2022/04/the-power-of-natural-language-processing
[Accessed 30 June 2023].

Jonathan P. Scaccia, V. C. S., 2021. 5335 days of Implementation Science: using natural language processing to examine publication trends and topics. *Implementation Science,* 16(47).

Le Glaz A, H. Y. K.-D. D. L. P. B. R. R. T. M. J. D. J. W. M. B. S. L. C., 2021. Machine Learning and Natural Language Processing in Mental Health: Systematic Revie. *J Med Internet Res,* 23(5), pp. 1438-8871.

Mah, P. & Skalna, I., 2022. Natural Language Processing and Artificial Intelligence for Enterprise Management in the Era of Industry 4.0. *Applied Sciences,* 12(9207), pp. 1-26.

Nadkarni, P. M., 2011. Natural language processing: an introduction. *Journal of the American Medical Informatics Association,* 18(5), pp. 544-551.

Ronan Collobert, J. W. L. B. M. K. K. K. P. K., 2000. Natural Language Processing (almost) from Scratch. *Journal of Machine Learning Research,* pp. 1-48.

Santosh K Behera, M. M. N., 2020. Natural Language Processing for Text and Speech Processing: A Review Paper. *International Journal of Advanced Research in Engineering and Technology (IJARET) ,* p. 6.

Sethunya R Joseph, H. H. K. L. F. K. K. S., 2016. Natural Language Processing: A Review. *International Journal of Research in Engineering and Applied Sciences,* 6(3), pp. 207-212.

van Erp M, R. C. M. D. S. A. I. M. R. A. F. L. M. A. d. T. D. S. R. X. T. C. B. S. A. M. C. K. A. F. A. B. S. L. R. R. F. T. d. S. J. a. B. U., 2021. Using Natural Language Processing and Artificial Intelligence to Explore the Nutrition and Sustainability of Recipes and Food. *Frontiers in Artificial Intelligence,* Volume 3, pp. 1-8.