5/7/24, 10:19 AM Labwork1

```
In [ ]: import nltk
                     import nitk
from nitk.stem import PorterStemmer, LancasterStemmer, SnowballStemmer, RegexpSter
from nitk.stem.wordnet import WordNetLemmatizer
from nitk.tokenize import word_tokenize
from nitk.corpus import stopwords
from os import path
from PIL import Image
                      from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
                     import matplotlib.pyplot as plt
                    paragraph1 = "You are musicophile," One who loves music". In your interactions with your MCA classmates you found another musicophile ,"Ameyatma"(classmate) and become friends. Now you want paragraph2 = "Social media mining is the process of obtaining big data from user-generated content on social media sites and mobile apps in order to extract actionable patterns, form conclus
                    ps = PorterStemmer()
                     lemmatizer = WordNetLemmatizer()
                   tokens1 = word_tokenize(paragraph1)
tokens2 = word_tokenize(paragraph2)
                    # Stop words removal for both paragraphs
                    stop words = set(stopwords.words('english'))
                     filtered_sentence1 = [w for w in tokens1 if w.lower() not in stop_words]
filtered_sentence2 = [w for w in tokens2 if w.lower() not in stop_words]
                   # Porter Stemming for both Paragraph
porter_sentence1 = [ps.stem(w) for w in filtered_sentence1]
porter_sentence2 = [ps.stem(w) for w in filtered_sentence2]
                    # Lemmatization for both sentences
lemmatized_tokens1 = [lemmatizer.lemmatize(word) for word in filtered_sentence1]
lemmatized_tokens2 = [lemmatizer.lemmatize(word) for word in filtered_sentence2]
                    max_words=100,
max_font_size=50,
                                                                                        random state=42
                                                                                   ).generate_from_text(" ".join(porter_sentence1))
                    print("Worcloud of Paragraph 1 after Stemming ")
fig = plt.figure(1)
plt.imshow(withstemmingwordcloud1)
                     plt.axis('off')
                     nlt.show()
                    max_words=100,
                                                                                      max font size=50,
                                                                                        random state=42
                                                                                   ).generate_from_text(" ".join(porter_sentence2))
                    print("Worcloud of Paragraph 2 after Stemming ")
fig = plt.figure(1)
plt.imshow(withstemmingwordcloud2)
plt.awis('off')
plt.show()
                     plt.show()
                    max words=100,
                                                                                      max font size=50,
                                                                                        random state=42
                                                                                   ).generate_from_text(" ".join(lemmatized_tokens1))
                    print("Morcloud of Paragraph 1 after Lammetization ")
fig = plt.figure(1)
plt.imshow(withLammetizationwordcloud1)
plt.axis('off')
                     plt.show()
                    max_words=100,
max_font_size=50,
                                                                                        random state=42
                                                                                   ).generate_from_text(" ".join(lemmatized_tokens2))
                   print("Worcloud of Paragraph 2 after Lammetization ")
fig = plt.figure(1)
plt.imshow(withLammetizationwordcloud2)
plt.axis('off')
plt.show()
                    after Stemming certain words in the word cloud are not complete but in lemmatization it has proper words this is seen in both the paragraphs
                  Worcloud of Paragraph 1 after Stemming
                  stanza wantexampl strang SON grip wantexample wantexam
                  Worcloud of Paragraph 2 after Stemming
```

5/7/24, 10:19 AM Labwork1





Worcloud of Paragraph 2 after Lammetization

