

```

In [ ]: 1 #Loading the packages
2 import os
3 import pandas as pd
4
5 #Importing the SBERT pre-trained-model
6 from sentence_transformers import SentenceTransformer
7 embedder = SentenceTransformer('all-mpnet-base-v2')
8
9 #Importing the kmeans clusterign and the Word Cloud package
10 from sklearn.cluster import KMeans
11 import matplotlib.pyplot as plt
12 from wordcloud import WordCloud
13
14 # Loading the data
15 df = pd.read_excel('Criteria_description.xlsx')
16 df.head()
17
18 #Making the Sentence embedding operations
19 corpus = list(df['Definition'])
20 corpus
21 corpus_embeddings = embedder.encode(corpus)
22 corpus_embeddings
23
24 #Making the kmeans clustering operations
25 clustering_model = KMeans(n_clusters=25, random_state=0, n_init=300)
26 clustering_model.fit(corpus_embeddings)
27 cluster_assignment = clustering_model.labels_
28 cluster_assignment
29
30 #Print to spreadsheet
31 cluster_df = pd.DataFrame(corpus, columns = ['corpus'])
32 cluster_df['cluster'] = cluster_assignment
33 cluster_df['code'] = df['code']
34 cluster_df.head()
35 file_name = 'Grouped_topics_thesis_bert.xlsx'
36 cluster_df.to_excel(file_name)
37 print
38
39 #Clusters word clouds
40 def word_cloud(pred_df,label):
41     wc = ''.join([text for text in pred_df['corpus'][pred_df['cluster'] == label]])
42     wordcloud = WordCloud(width=800, height=500,
43         random_state=21, max_font_size=110).generate(wc)
44     fig7 = plt.figure(figsize=(10, 7))
45     plt.imshow(wordcloud, interpolation="bilinear")
46     plt.axis('off')
47
48 #For each cluster replace 0 to (1, ..., 24)
49 word_cloud(cluster_df,0)

```

```

In [ ]: 1

```