

```

1  __author__ = 'HITESH'
2  from PIL import Image
3  from os.path import dirname, isfile, join
4  from os import listdir
5  from wordcloud import WordCloud
6  import matplotlib.pyplot as plt
7  import numpy as np
8
9  def getFinalUGCollName(collName):
10     if collName.find("IIT") != -1:
11         return "IIT"
12     if collName.find("NIT") != -1:
13         return "NIT"
14     if collName.find("BITS") != -1:
15         return "BITS"
16     collName = collName.replace(" ", "").replace("of", "").replace("of", "").replace(".", "").replace(",", "").replace("&", "")
17     collName = collName.replace("-", "").replace("'", "").replace("Information", "Info").replace("information", "info")
18     collName = collName.replace("university", "Univ").replace("University", "Univ").replace("Institute", "Inst")
19     collName = collName.replace("Institute", "Inst").replace("College", "Coll").replace("college", "Coll")
20     collName = collName.replace("Engineering", "Engg").replace("engineering", "Engg").replace("Technology", "Tech")
21     collName = collName.replace("technology", "Tech").replace("Science", "Sci").replace("science", "Sci")
22     return collName
23
24 #generate input for wordle generator
25 def generateUnivVsUGCollegeMapForWordleGenerator():
26     univVsUGCollegeMap = {}
27     for entry in open(dirname(__file__) + '/../scraper/univJSON/allUsersFinal typo correct.csv'):
28         if entry.startswith("userName"):
29             continue
30         parts = entry.split(",")
31         univ = parts[24]
32         ugCollege = parts[18]
33
34         if univ in univVsUGCollegeMap:
35             if ugCollege.strip() != "":
36                 univVsUGCollegeMap[univ].append(ugCollege)
37         else:

```

```

38             univVsUGCollegeMap[univ] = [ugCollege]
39     return univVsUGCollegeMap
40
41 #create input files from the map
42 def createInputFilesForWordleGenerator(univVsUGCollegeMap):
43     for univ in univVsUGCollegeMap:
44         outputFile = open("input/" + univ.replace("/", " ") + ".txt", "w")
45         ugCollegeArray = univVsUGCollegeMap[univ]
46         ugCollegeArray = [getFinalUGCollName(ugCollege) for ugCollege in ugCollegeArray]
47         outputFile.write('\n'.join(ugCollegeArray))
48         outputFile.close()
49
50 def generateWordleForInput(textFile):
51     text = open("input/" + textFile).read()
52     wordcloud = WordCloud().generate(text)
53     '''image = wordcloud.to_image()
54     image.show()'''
55
56     '''plt.imshow(wordcloud)
57     plt.axis("off")
58     plt.show()'''
59
60     wordcloud = WordCloud(min_font_size=10, max_font_size=60, relative_scaling=.5).generate(text)
61     plt.figure()
62     plt.imshow(wordcloud)
63     plt.axis("off")
64     plt.savefig("output/" + textFile + ".png")
65     plt.close()
66
67     #mask
68     '''wordCloud = WordCloud(background_color="white", max_words=2000, mask=msMask)
69     wordCloud.generate(text)
70     wordCloud.to_file("output/" + textFile.replace(".txt", "") + ".png")'''
71
72 '''univVsUGCollegeMap = generateUnivVsUGCollegeMapForWordleGenerator()
73 createInputFilesForWordleGenerator(univVsUGCollegeMap)'''
74 #msMask = np.array(Image.open("mask.png"))

```

```
75
76 '''inputFiles = [f for f in listdir("input/") if isfile(join("input/", f))]
77 for inputFile in inputFiles:
78     if ".txt" in inputFile:
79         generateWordleForInput(inputFile)
80         print inputFile, " done"'''
81 generateWordleForInput("allUGCollegesList.txt")
```