import pandas as pd

import warnings

warnings.filterwarnings(action = 'ignore', category = UserWarning, module = 'gensim')

from gensim.models.keyedvectors import KeyedVectors

import time

import random

from pymongo import MongoClient

a = input("請輸入關鍵字：")

def env\_search(userstr):

word = pd.read\_excel('all.xlsx')

url = word['網址']

word1=word["內文"]

title=word["標題"]

dict\_1 = ["約會", "上課", "上班", "旅遊", "聚會", "出門","霧眉"]

dict\_2 = ["韓系", "日系", "歐美", "眼妝", "口紅", "修容", "全妝", "唇膏", "淡妝", "濃妝", "妝容"]

Index\_1= [] #第一層相關字的index

str\_1 = "" #第一層相關字

Index\_2= [] #第二層相關字的index

str\_2= "" #第二層相關字

Index\_3= [] #所得的結果

#第一層

for i in dict\_1:

if i in userstr:

str\_1= i

for n, article in enumerate(word1):

if str\_1 in str(article):

Index\_1.append(n)

else:

break

# 第二層

for i in dict\_2:

if i in userstr :

str\_2= i

for s in Index\_1:

if str\_2 in str(word1[s]):

Index\_2.append(s)

else:

break

if Index\_2:

Index\_3 = Index\_2

else:

Index\_3 = Index\_1

if Index\_3:

Index\_3=Index\_3

else:

result="無相關資料"

return result

# 利用work2vec比對詞向量

score = []

seg = [] #文章的index和詞向量分數

if str\_2:

str\_2=str\_2

else:

str\_2= "妝"

text1=str\_1+str\_2

print("text1: " ,text1)

print("Index3 : " ,Index\_3)

print("title:" ,title)

for i in Index\_3:

try:

similar = model.wv.wmdistance(text1, title[i])

com = (i, similar,title[i])

seg.append(com)

score.append(similar)

except:

pass

print("seg :" ,seg)

# 排序文章分數

sort\_score=[]

ans = sorted(score)

for i in ans:

if i not in sort\_score:

sort\_score.append(i)

# 選出三則貼文

highest=[]

for n in range(0,len(seg)):

for y in sort\_score[0:5]: #選擇前5高的分數

if seg[n][1] == y:

highest.append(seg[n][0])

if len(highest) > 3:

final = random.sample(highest, 3)

elif len(highest) ==0:

result = "無相關資料"

return result

else:

final=highest

print(final)

print(sort\_score)

# # 連接mongo提取文章資料

# mongo\_client=MongoClient('mongodb://10.120.38.27:15000/')

# db=mongo\_client.env\_search

# result=[] #最終結果

# for i in final:

# tag\_ = go + "X" + gogo

# result.append(tag\_)

# mongodb=db.segment.find({'\_id':i})

# mongoresult=""

# for x in mongodb:

# mongoresult=x

# for key,value in mongoresult.items():

# if key == "標題" or key =="網址":

# result.append(value)

# if result:

# result=resultㄑ

# else:

# result="無相關資料"

# return result

env\_search(a)