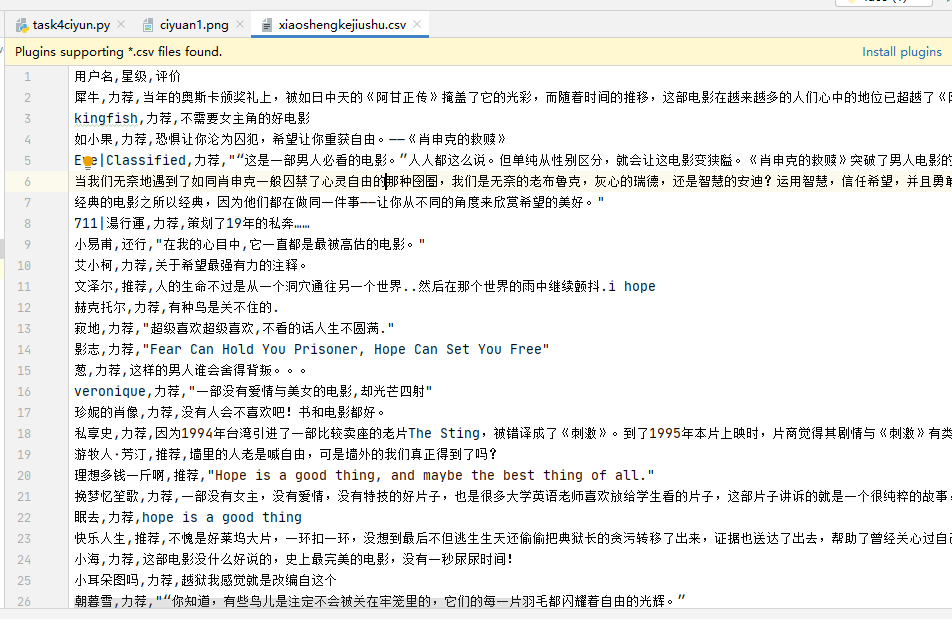
4. 通过网络爬虫爬取网上 <https://movie.douban.com/subject/1292052/comments?status=P>

。肖申克的救赎 短评前100条评论，爬取信息为：用户名，星级，评价。将爬取的信息存储到本地的CSV文件中。（80分）然后制作成词云（20分）

代码：

import urllib.request  
import random  
import csv  
from bs4 import BeautifulSoup  
import numpy  
import jieba  
from PIL import Image  
from wordcloud import WordCloud  
  
*# 读取肖申克救赎  
# 列表里放列表*csvlist = []  
*# 获取头*def getHeader(url):  
 *#谷歌* header1={  
 **'Host'**:**'movie.douban.com'**,  
 **'User-Agent'**:**'Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/77.0.3865.120 Safari/537.36'** }  
 *#IE* header2={  
 **'Host'**: **'movie.douban.com'**,  
 **'User-Agent'**: **'Mozilla/5.0 (Windows NT 6.1; WOW64; Trident/7.0; rv:11.0) like Gecko'** }  
 *# 获取请求* list = [header1, header2]  
 index = random.choice(list)  
 req = urllib.request.Request(url=url, headers=index)  
 return req  
  
*# 获取数据*def getData(url):  
 req = getHeader(url)  
 html = urllib.request.urlopen(req)  
 data = html.read()  
 *#print(data)* soup = BeautifulSoup(data, **'html.parser'**)  
 *#print(soup)  
 # 1.获取一个大层div* commentsdiv = soup.select(**'.mod-bd'**)[0]  
 *# print(commentsdiv)  
 # 2.获取评论div* commentsitem = commentsdiv.select(**'.comment-item'**)  
 *# 3.循环评论div* for i in commentsitem:  
 *# 先获取用户名和星级div* uapdiv = i.select(**'.comment-info'**)[0]  
 *# 3.1.用户名* username = uapdiv.select(**'a'**)[0].string  
 *# 3.2 星级* starspan = uapdiv.select(**'.rating'**)[0]  
 star = starspan.get(**'title'**)  
 *# 3.3 评价* comment = i.select(**'.short'**)[0].string  
 *# 4 组成data列表添加到list中* datas = [username, star, comment]  
 csvlist.append(datas)  
 *# print(username)  
 # print(star)  
 # print(comment))  
  
#csv操作*def csvf():  
 with open(**'xiaoshengkejiushu.csv'**, **'w'**, newline=**''**, encoding=**'utf-8'**) as openfile:  
 f = csv.writer(openfile)  
 data = [**'用户名'**,**'星级'**,**'评价'**]  
 f.writerow(data)  
 *# 循环写入* for i in csvlist:  
 f.writerow(i)  
  
*# 制作词云  
# 1.获取csv的评论数据*def getGoodCmtList():  
 str = **""** with open(**'xiaoshengkejiushu.csv'**, **'r'**, newline=**''**, encoding=**'utf-8'**) as openfile:  
 data = csv.reader(openfile)  
 for i in data:  
 *# 评论放进去* str += i[2]  
 return str  
  
*# 2.具体实施词云*def generwordcloud():  
 *#2.1获取csv字符串数据* str = getGoodCmtList()  
 *#2.1获取关键文字* cutCloud = **"??"**.join(jieba.cut(str))  
 *# print(cutCloud)  
 #2.2读取图片* bgImg = numpy.array(Image.open(**'cloudimg/t2.jpg'**))  
 *#2.3生成词云对象,选择本地字体，背景图片，背景颜色* cloud = WordCloud(  
 font\_path=**"C:\Windows\Fonts\ARIALUNI.TTF"**,  
 mask=bgImg,  
 background\_color=**"pink"** ).generate(cutCloud)  
 *#2.4生成保存词云图片* cloud.to\_file(**"ciyuan1.png"**)  
  
  
*# 入口函数*if \_\_name\_\_ == **'\_\_main\_\_'**:  
 *# 获取数据,前100的话，就是获取5遍，每次获取20个评论* start = 0  
 for i in range(0, 5):  
 getData(**'https://movie.douban.com/subject/1292052/comments?start=%d&limit=20&status=P&sort=new\_score'**%start);  
 start += 20  
 print(**'100条评论读取完成'**)  
 *# 写入数据* csvf()  
 *# 制作词云* generwordcloud()  
 print(**'词云图片制作完成'**)

截图：





5.使用 opencv对下面图片进行人脸识别。（80分）

代码：

import cv2  
if \_\_name\_\_ == **'\_\_main\_\_'**:  
 *#1.获取原图片* img = cv2.imread(**'opencvimg/nvtuan.jpg'**)  
 *#2.导入并且读取人脸特征数据包* faceData = cv2.CascadeClassifier(**'haarcascade\_frontalface\_alt.xml'**)  
 *#3.人脸数据匹配,图片源，扫描系数，左右匹对* face = faceData.detectMultiScale(img, scaleFactor=1.05, minNeighbors=7)  
 for x,y,w,h in face:  
 *#4.对人脸范围画宽* cv2.rectangle(img, pt1=(x,y), pt2=(x+w,y+h), color=[0,0,255],thickness=2)  
 *#5.展示窗口* cv2.imshow(**"win"**,img)  
 *#6.不要关闭* cv2.waitKey()  
 *#7.回收资源* cv2.destroyWindow(**"win"**)

截图：

