作业一：

该代码实现的是对网站<https://pvp.qq.com/web201605/herolist.shtml>的数据爬取，主要包括两个部分，一个是数据采集，一个是数据入库

MysqlClient类，是实现的是mysql数据库操作，封装了3个函数，save()实现的是对传入的sql语句执行增加删除修改的execute操作，find\_all（）实现的是对传入的sql语句的查找操作，实现的是fetchall方法，find\_one（）实现的是对传入的sql语句的查找操作，这个是查找某一个元素，实现的是fetchone方法。

数据采集包括了get\_weapon()，get\_jineng()，get\_hero()，get\_hero\_detail()，get\_video()，5个函数，这个5个函数的功能都是大同小异，用的都是request库去请求网页数据，不同的内容请求的url地址不一样，比如获取英雄数据，请求的url是https://pvp.qq.com/web201605/js/herolist.json，对得到的返回数据用json进行解析，获取相应的字段，用MysqlClient类实例化的对象，执行对应的sql语句，进行存储到mysql数据库操作。

作业二：

问题 a)：

方法一：在当前目录添加 alice\_mask.png图片，stopwords.txt停用词文件，simhei.ttf字体文件。

方法二：在当前目录添加 alice\_mask.png图片，把原文中对应的80到82行代码注释掉，即不使用停用词（但是这样生成的词云会有较多无意义的词）。然后原文中的93行对应的字体文件使用绝对路径，比如修改为：C:\Windows\Fonts\SimHei.ttf。

问题 b)：

方法一：

将item.find\_all('p')[0].string 修改为item.find\_all('p')[0].get\_text()

方法二：

将item.find\_all('p')[0].string 修改为item.find\_all('span',class\_='short')[0].get\_text()

问题 c)：

对应的51-53行修改为：

for item in comment\_div\_lits:  
 try:  
 if item.find\_all('p')[0].get\_text() is not None:  
 eachCommentList.append(item.find\_all('p')[0].get\_text())  
 except:  
 continue

python shell上的运行结果为：



生成的词云图为：

