

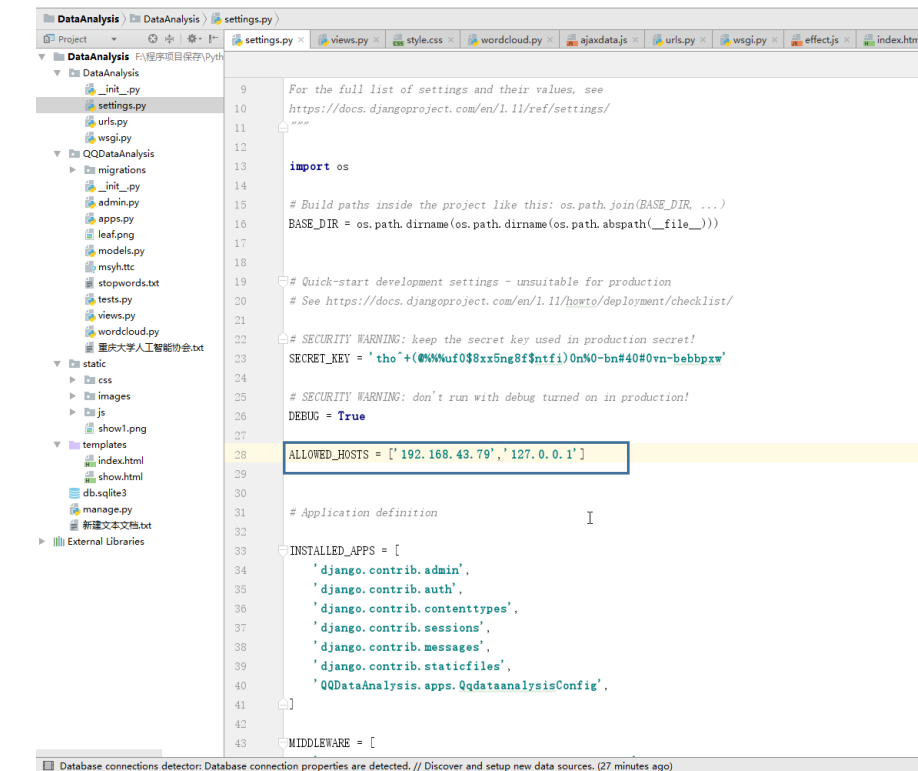
WordCloud 使用说明

1. 项目简介：

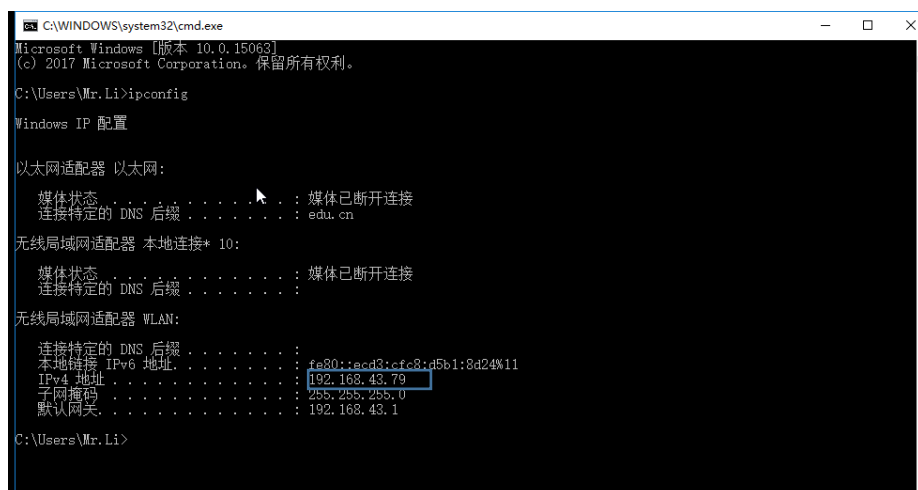
通过 Python 的 Django 框架搭建了一个网站，并把词云图生成的 python 代码内嵌在了这个网站里面。

2. 项目演示：

- (1) 用打开 DataAnalysis 项目，打开 settings.py 文件，在 ALLOWED_HOSTS 列表里加入本机在局域网中的 ip 地址，然后更改 index.html 文件的代码，改成你自己电脑的 ip 地址（注意：只改 ip 地址）



```
9 For the full list of settings and their values, see
10 https://docs.djangoproject.com/en/1.11/ref/settings/
11
12
13 import os
14
15 # Build paths inside the project like this: os.path.join(BASE_DIR, ...)
16 BASE_DIR = os.path.dirname(os.path.dirname(os.path.abspath(__file__)))
17
18
19 # Quick-start development settings - unsuitable for production
20 # See https://docs.djangoproject.com/en/1.11/howto/deployment/checklist/
21
22 # SECURITY WARNING: keep the secret key used in production secret!
23 SECRET_KEY = 'tho'+(0%Wuf0$8xx5ng8f$ntfi)0nW0-bn#40#0vn-bebbpxw'
24
25 # SECURITY WARNING: don't run with debug turned on in production!
26 DEBUG = True
27
28 ALLOWED_HOSTS = ['192.168.43.79', '127.0.0.1']
29
30
31 # Application definition
32
33 INSTALLED_APPS = [
34     'django.contrib.admin',
35     'django.contrib.auth',
36     'django.contrib.contenttypes',
37     'django.contrib.sessions',
38     'django.contrib.messages',
39     'django.contrib.staticfiles',
40     'QQDataAnalysis.apps.QqdataanalysisConfig',
41 ]
42
43 MIDDLEWARE = [
```



```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [版本 10.0.15063]
(c) 2017 Microsoft Corporation。保留所有权利。

C:\Users\Mr.Li>ipconfig

Windows IP 配置

以太网适配器 以太网:

    媒体状态 . . . . . : 媒体已断开连接
    连接特定的 DNS 后缀 . . . . . : edu.cn

无线局域网适配器 本地连接* 10:

    媒体状态 . . . . . : 媒体已断开连接
    连接特定的 DNS 后缀 . . . . . :

无线局域网适配器 WLAN:

    连接特定的 DNS 后缀 . . . . . :
    本地链接 IPv6 地址 . . . . . : fe80::ecd3:cfc8:d5b1:8d24%11
    IPv4 地址 . . . . . : 192.168.43.79
    子网掩码 . . . . . : 255.255.255.0
    默认网关 . . . . . : 192.168.43.1

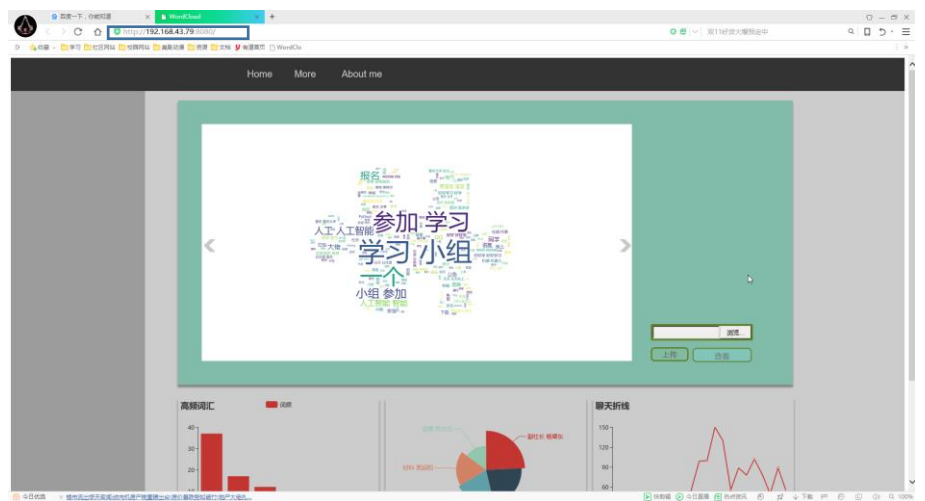
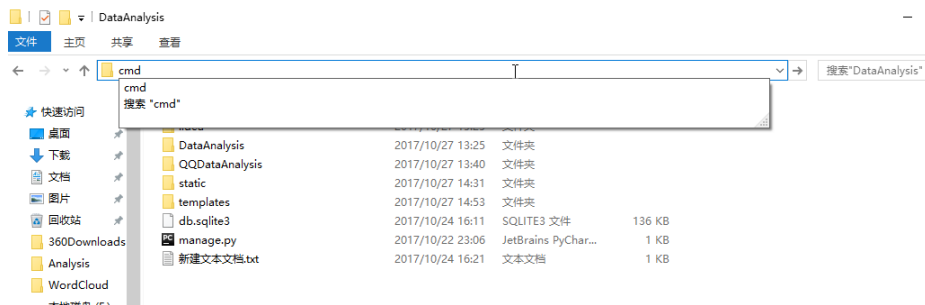
C:\Users\Mr.Li>
```

```

<div id="word-cloud">
  <div id="arrow-container">
    <div id="pre" class="arrow"></div>
    <div id="next" class="arrow"></div>
  </div>
  <div id="image-container">
    <div id="img-1" class="img"></div>
    <div id="img-2" class="img"></div>
    <div id="img-3" class="img"></div>
    <div id="img-4" class="img"></div>
    <div id="img-5" class="img"></div>
  </div>
  <div id="select-container">
    <p>
      <a href="http://192.168.43.79:8080/check" id="show" target="_blank">查看</a>
    </p>
    <form>
      <input type="file" name="file" id="select-file">
      <input type="submit" value="上传" id="submit-btn">
    </form>
  </div>
  <div id="chart-1" class="chart"></div>
  <div id="chart-2" class="chart"></div>
  <div id="chart-3" class="chart"></div>
</div>

```

- (2) 退出项目，在该项目的目录下（有 manage.py 文件）打开 cmd，输入命令行：
python manage.py runserver 0.0.0.0:8080 回车，启动服务器。通过后，局域网里所有的计算机输入服务器的 ip 地址加 8080 端口号都可以访问该网站了



网页中有几张事先生成的词云图和词频统计表。

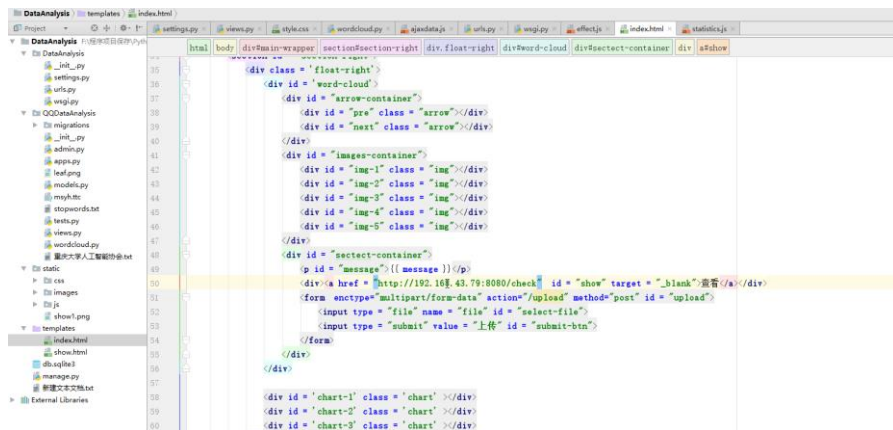
- (3) 点击浏览，选择聊天记录，上传。选择框上如果显示 Get File，证明聊天记录上传成功。然后点击查看，就会在新窗口把词云图显示出来，并在下面显示词频统计的结果。

P.S.这个过程可能有点漫长，因为我算法优化的原因，词云图生成比较慢，而且这也和选择做服务器的电脑的性能也有关。



3. 可能存在的问题：

- (1) 端口被占用：如果你的电脑上 8080 端口不可使用，或者你想换端口，就在启动服务器的时候输入 runserver 0.0.0.0:其他端口号。然后在代码里面的 index.html 改端口号



- (2) 在上文件的时候出 bug。我的这个网站上传的文件只支持 QQ 导出的历史消息（因为代码是针对 QQ 历史消息的格式设计的）。而且 txt 文件的编码格式是 UTF-8，不然无法生成词云图。
- (3) 网站首页的词云图幻灯片有 bug（不能实现图片轮播），但不影响词云图的生成
- (4) 如果学长无法正常测试我的代码的话，可以测试那个纯 python 的代码（Analysis 项目），那个代码就没有可视化界面，也不能生成图表。但能正常生成云词图，和词频统计
- (5) 在生成第一个统计列表的时候，存在一个 bug，但不影响生成，其实可以修复，但会导致算法的复杂度变大（本来算法的复杂度就很大）。

4. 总结

这次的项目由我一人独立完成，还是付出了不少汗水吧。在做项目之前，就只了解过 Python 的基础知识和一点网站前端的知识，什么 Django, json, Ajax 都是现学现卖的。因为时间有点紧，还有一些功能和美化没有实现，但是我也算是完成任务了。我是本着兴趣参加这次骨干选拔的，我知道，我大一时错过了很多的活动，也错过了很多机会，荒废了一年的时间。所以这次的机会我格外重视，认真完成一个项目后，收获还是很大的。所以我很想加入创新实验班，和更多志同道合的高手们交流，去做更多好玩有趣的项目。