Python编程新思维及实战

嵩 天

实例2：四大名著词云分析

嵩天

Python编程新思维及实战

"四大名著词云分析"需求分析

Python123

需求分析

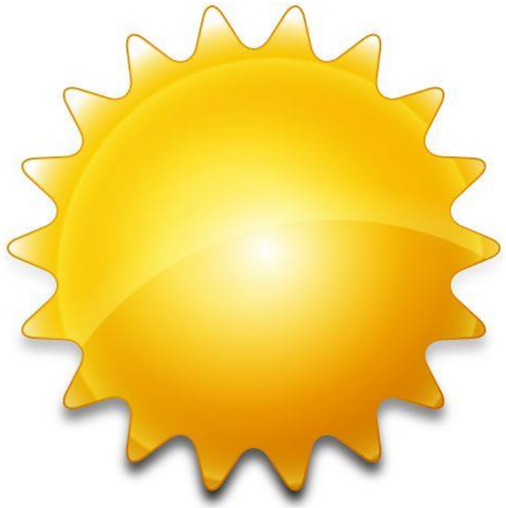
中国古典四大名著的词云效果

- 输入：四大名著的txt文本文件
- 输出：词云效果及有形状的词云

需求分析

中国古典四大名著的词云效果

长方形词云



Python编程新思维及实战

"四大名著词云分析"实例编写

Python123

代码概览：传统版本

```
# for WordCloud
import jieba
import wordcloud
names = {"红楼梦.txt", "三国演义.txt", "水浒传.txt", "西游记.txt"}

for name in names:
    f = open(name, "r", encoding="utf-8")
    t = f.read()
    f.close()
    ls = jieba.lcut(t)
    txt = " ".join(ls)
    w = wordcloud.WordCloud(   font_path = "msyh.ttc",\
            width = 1000, height = 700 \
    )
    w.generate(txt)
    w.to_file(name.split(".")[0] + ".png")
```

代码详解

```
# for WordCloud
import jieba
import wordcloud
names = {"红楼梦.txt", "三国演义.txt", "水浒传.txt", "西游记.txt"}

for name in names:
    f = open(name, "r", encoding="utf-8")
    t = f.read()
    f.close()
    ls = jieba.lcut(t)
    txt = " ".join(ls)
    w = wordcloud.WordCloud(   font_path = "msyh.ttc",\
            width = 1000, height = 700 \
    )
    w.generate(txt)
    w.to_file(name.split(".")[0] + ".png")
```

引入中文分词和词云库

代码详解

```
# for WordCloud
import jieba
import wordcloud
names = {"红楼梦.txt", "三国演义.txt", "水浒传.txt", "西游记.txt"}

for name in names:
    f = open(name, "r", encoding="utf-8")
    t = f.read()
    f.close()
    ls = jieba.lcut(t)
    txt = " ".join(ls)
    w = wordcloud.WordCloud(   font_path = "msyh.ttc",\
            width = 1000, height = 700 \
    )
    w.generate(txt)
    w.to_file(name.split(".")[0] + ".png")
```

遍历四大名著的txt文件

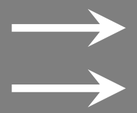
代码详解

```
# for WordCloud
import jieba
import wordcloud
names = {"红楼梦.txt", "三国演义.txt", "水浒传.txt", "西游记.txt"}

for name in names:
    f = open(name, "r", encoding="utf-8")
    t = f.read()
    f.close()
    ls = jieba.lcut(t)
    txt = " ".join(ls)
    w = wordcloud.WordCloud(   font_path = "msyh.ttc",\
            width = 1000, height = 700 \
    )
    w.generate(txt)
    w.to_file(name.split(".")[0] + ".png")
```

读入名著，读取全部内容

代码详解

```
# for WordCloud
import jieba
import wordcloud
names = {"红楼梦.txt", "三国演义.txt", "水浒传.txt", "西游记.txt"}

for name in names:
    f = open(name, "r", encoding="utf-8")
    t = f.read()
    f.close()
    ls = jieba.lcut(t)
    txt = " ".join(ls)
    w = wordcloud.WordCloud(   font_path = "msyh.ttc",\
            width = 1000, height = 700 \
    )
    w.generate(txt)
    w.to_file(name.split(".")[0] + ".png")
```



按照wordcloud库要求分词后以空格分隔再组合

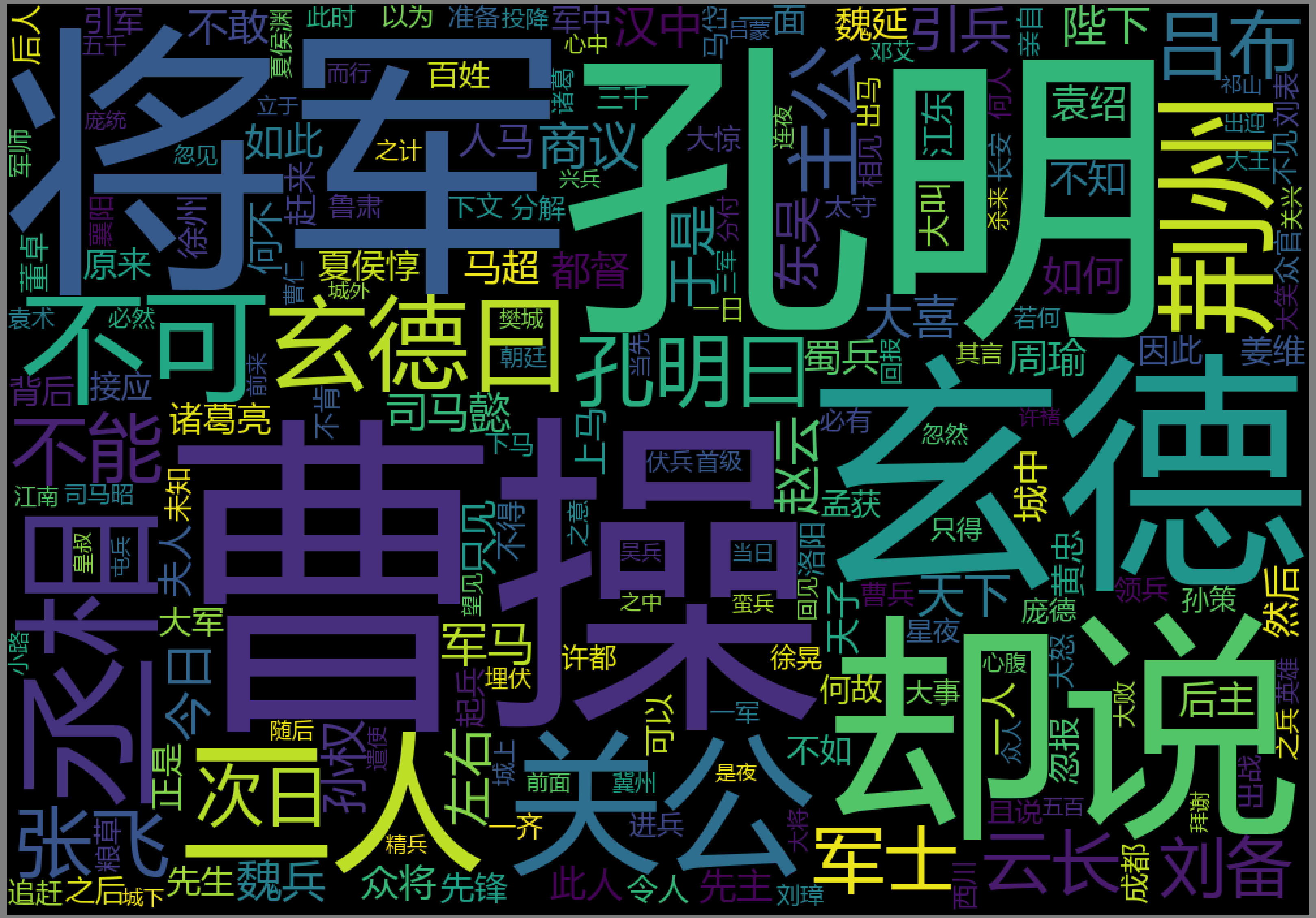
代码详解

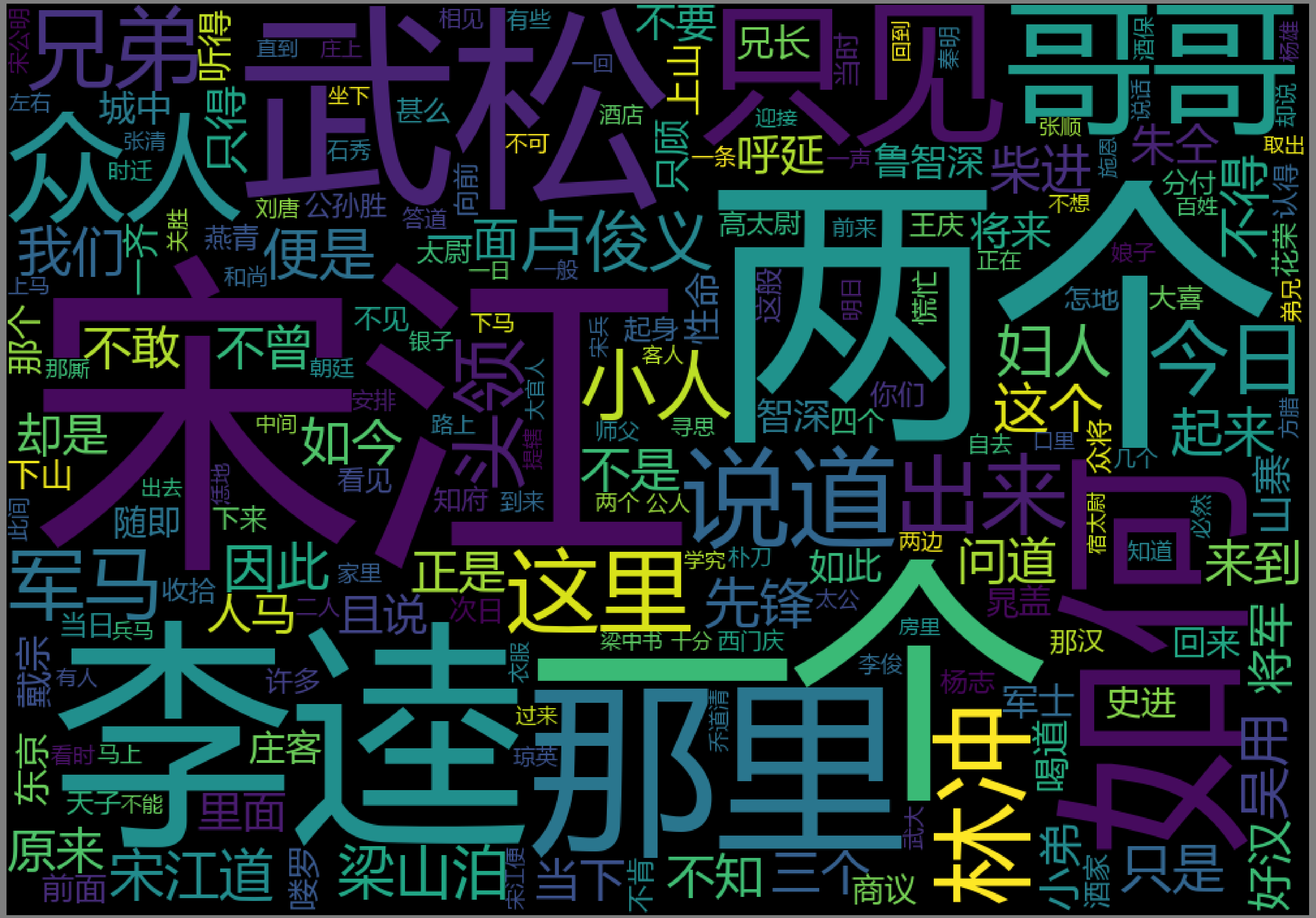
```
# for WordCloud
import jieba
import wordcloud
names = {"红楼梦.txt", "三国演义.txt", "水浒传.txt", "西游记.txt"}

for name in names:
    f = open(name, "r", encoding="utf-8")
    t = f.read()
    f.close()
    ls = jieba.lcut(t)
    txt = " ".join(ls)
    w = wordcloud.WordCloud(   font_path = "msyh.ttc",\
            width = 1000, height = 700 \
    )
    w.generate(txt)
    w.to_file(name.split(".")[0] + ".png")
```

设定微软雅黑字体

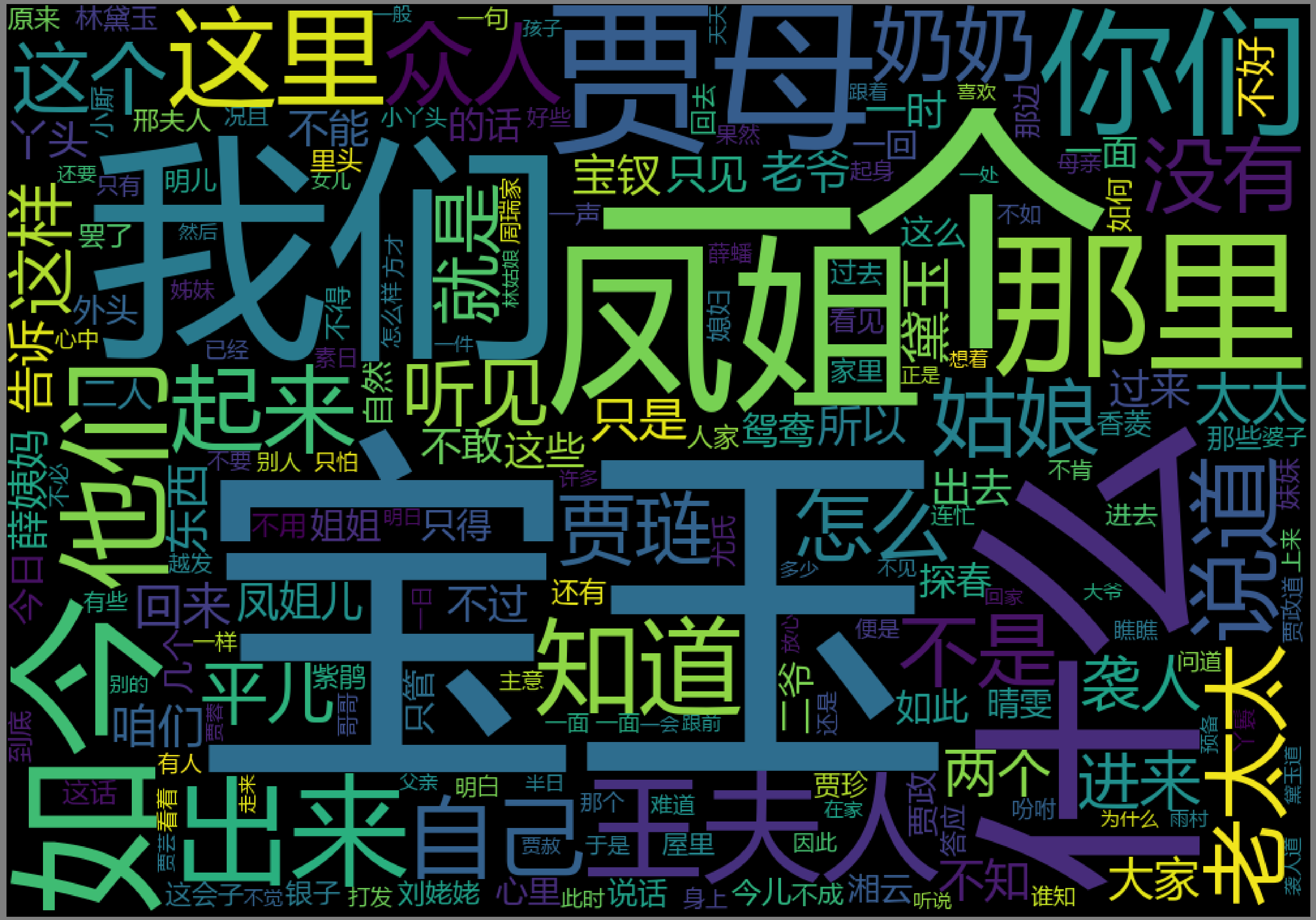
词云三部曲

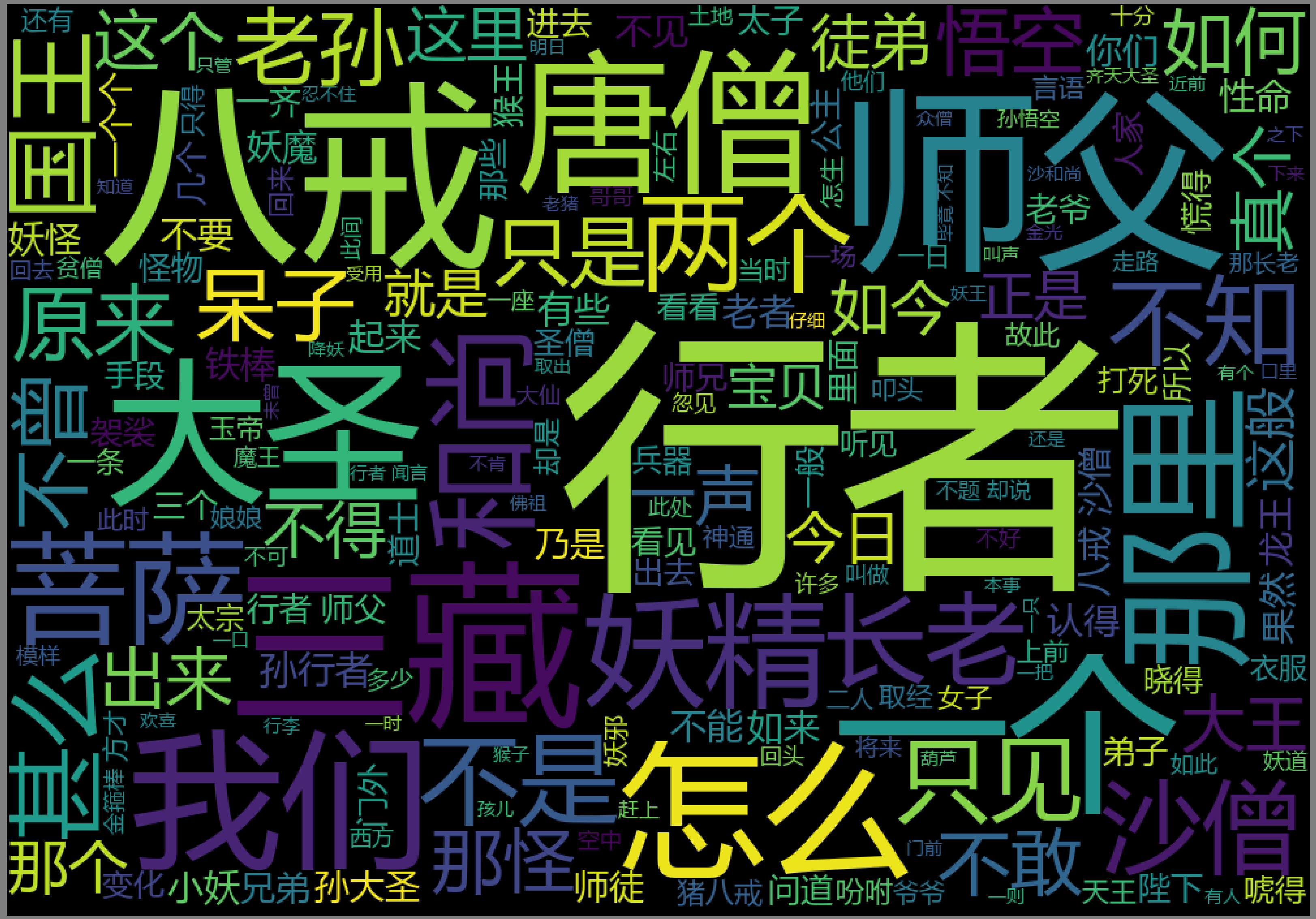
运行结果


曹操
孔明
将军
玄德
却说
关公
丞相
荆州
玄德曰
孔明曰
不可
张飞
吕布
二人
次日
军士
云长
刘备
主公


宋江
李逵
两个
一个
那里
武松
只见
哥哥
如何
林冲
众人
兄弟
说道
出来
卢俊义
头领
这里

运行结果


宝玉
我们
凤姐
一个
那里
贾母
奶奶
你们
这里
众人
太太
如今
他们
出来
起来
知道
王夫人
老太太
说道
姑娘
听见
就是
没有
这个
告诉
这样
袭人
贾琏
怎么
自己
进来
平儿
探春
不是
林黛玉
原来


行者
八戒
唐僧
师父
大圣
三藏
和尚
菩萨
妖精
长老
一个
我们
怎么
那里
沙僧
只见
两个
只是
老孙
这里
这个
悟空
如何
国王
原来
呆子
甚么
不知
不是
那怪

代码概览：小太阳版本

```
# for WordCloud
import jieba
import wordcloud
from scipy.misc import imread
mask = imread("sun.png")
names = {"红楼梦.txt", "三国演义.txt", "水浒传.txt", "西游记.txt"}
for name in names:
    f = open(name, "r", encoding="utf-8")
    t = f.read()
    f.close()
    ls = jieba.lcut(t)
    txt = " ".join(ls)
    w = wordcloud.WordCloud(   font_path = "msyh.ttc",\
            width = 1000, height = 700, mask = mask\
    )
    w.generate(txt)
    w.to_file(name.split(".")[0] + "2.png")
```

代码详解

```
# for WordCloud
import jieba
import wordcloud
from scipy.misc import imread
mask = imread("sun.png")
names = {"红楼梦.txt", "三国演义.txt", "水浒传.txt", "西游记.txt"}
for name in names:
    f = open(name, "r", encoding="utf-8")
    t = f.read()
    f.close()
    ls = jieba.lcut(t)
    txt = " ".join(ls)
    w = wordcloud.WordCloud(   font_path = "msyh.ttc",\
            width = 1000, height = 700, mask = mask\
    )
    w.generate(txt)
    w.to_file(name.split(".")[0] + "2.png")
```

引入图片

代码详解

```
# for WordCloud
import jieba
import wordcloud
from scipy.misc import imread
mask = imread("sun.png")
names = {"红楼梦.txt", "三国演义.txt", "水浒传.txt", "西游记.txt"}
for name in names:
    f = open(name, "r", encoding="utf-8")
    t = f.read()
    f.close()
    ls = jieba.lcut(t)
    txt = " ".join(ls)
    w = wordcloud.WordCloud(   font_path = "msyh.ttc",\
            width = 1000, height = 700, mask = mask\
    )
    w.generate(txt)
    w.to_file(name.split(".")[0] + "2.png")
```

→

设置图片

代码详解

```
# for WordCloud
import jieba
import wordcloud
from scipy.misc import imread
mask = imread("sun.png")
names = {"红楼梦.txt", "三国演义.txt", "水浒传.txt", "西游记.txt"}
for name in names:
    f = open(name, "r", encoding="utf-8")
    t = f.read()
    f.close()
    ls = jieba.lcut(t)
    txt = " ".join(ls)
    w = wordcloud.WordCloud(   font_path = "msyh.ttc",\
            width = 1000, height = 700, mask = mask\
    )
    w.generate(txt)
    w.to_file(name.split(".")[0] + "2.png")
```

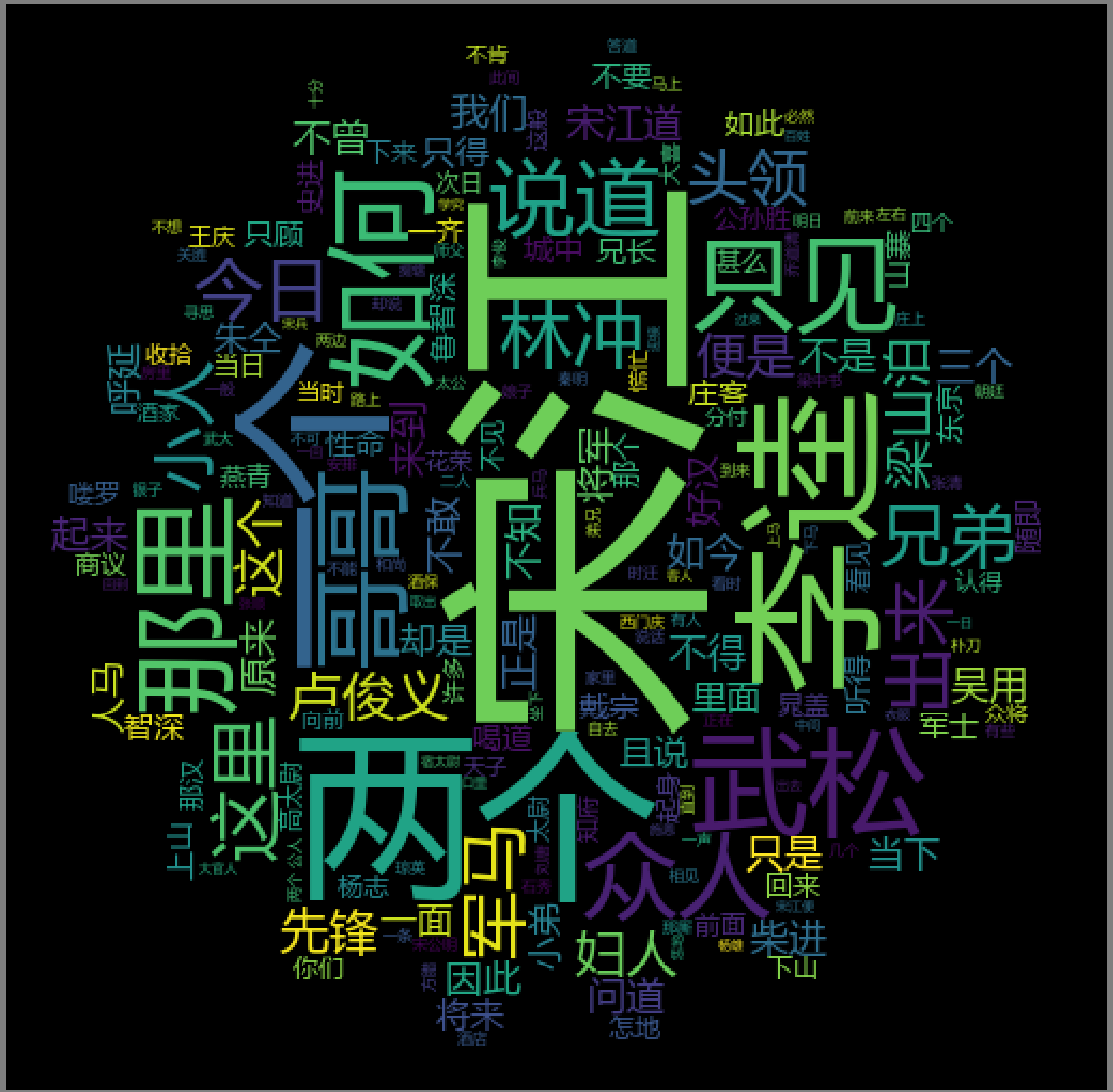
词云改名字

运行结果


曹操
孔明
将军
玄德
关公
却说
丞相
荆州
玄德曰
张飞
刘备
孔明曰
不可
吕布
东吴
赵云
都督
天下
引兵
商议
主公
军士
左右
次日
今日
陛下
于是
司马懿
孙权
魏延
军马
周瑜
袁绍
汉中
蜀兵
如何
云长
不能
先主
只见
此人
夫人
人马
马超


宝玉
什么
一个
我们
贾母
凤姐
那里
如今
你们
知道
众人
说道
王夫人
姑娘
奶奶
黛玉
这里
他们
自己
老太太
起来
出来
贾琏
平儿
袭人
两个
听见
所以
就是
咱们
告诉
太太
进来
老爷
回来
这样
东西
没有
大家
出去
宝钗
只是
凤姐儿
这个
怎么

运行结果



宋江
李逵
两个
武松
哥哥
如何
那里
说道
只见
林冲
众人
军马
兄弟
卢俊义
吴用
头领
妇人
先锋
小人
这里
出来

行者
师父
八戒
三藏
唐僧
大圣
一个
那里
和尚
怎么
妖精
我们
菩萨
沙僧
两个
不知
长老
老孙
国王
悟空
妖怪


Python 123
Thank you