**实战一：手把手教你中英文词云**

**1.简介**

词云是一种数据呈现方式。

**2.准备**

工具：python 3.6 sublime text

所需安装包:

pip install wordcloud matplotlib jieba pillow

**3.WordCloud()可选的参数**

font\_path：可用于指定字体路径，包括otf和ttf

width：词云的宽度，默认为400

height：词云的高度，默认为200

mask：蒙版，可用于定制词云的形状

min\_font\_size：最小字号，默认为4

max\_font\_size：最大字号，默认为词云的高度

max\_words：词的最大数量，默认为200

stopwords：将被忽略的停用词，如果不指定则使用默认的停用词词库

background\_color：背景颜色，默认为black

mode：默认为RGB模式，如果为RGBA模式且background\_color设为None，则背景将透明

**4.英文词云**

# \_\*\_ coding:utf-8 \_\*\_

# 进行英文词云生成

from wordcloud import WordCloud

import matplotlib.pyplot as plt

path = 'D://BaiduNetdiskDownload//深度有趣人工智能//实战一 词云的使用//'

# 打开文本

text = open(path + 'source//constitution.txt').read()

# 生成对象

wc = WordCloud().generate(text)

# 显示词云

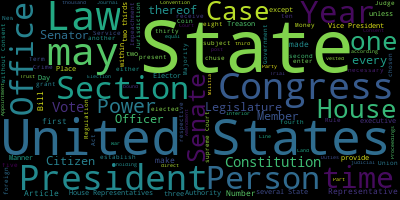
plt.imshow(wc,interpolation='bilinear')

plt.axis('off')

plt.show()

# 保存到文件

wc.to\_file(path + 'create\_images//wordcloud1.png')



由于英文单词之间有空格，所以不需要额外的处理。

**5.中文分词**

**（1）不分词的效果**

这里以西游记为例

# \_\*\_ coding:utf-8 \_\*\_

# 中文不分词效果展示

from wordcloud import WordCloud

import matplotlib.pyplot as plt

path = 'D://BaiduNetdiskDownload//深度有趣人工智能//实战一 词云的使用//'

# 打开文本

text = open(path + 'source//xyj.txt',encoding = 'UTF-8').read()

# 生成对象

wc = WordCloud(font\_path='Hiragino.ttf', width=800, height=600, mode='RGBA', background\_color=None).generate(text)

# 显示词云

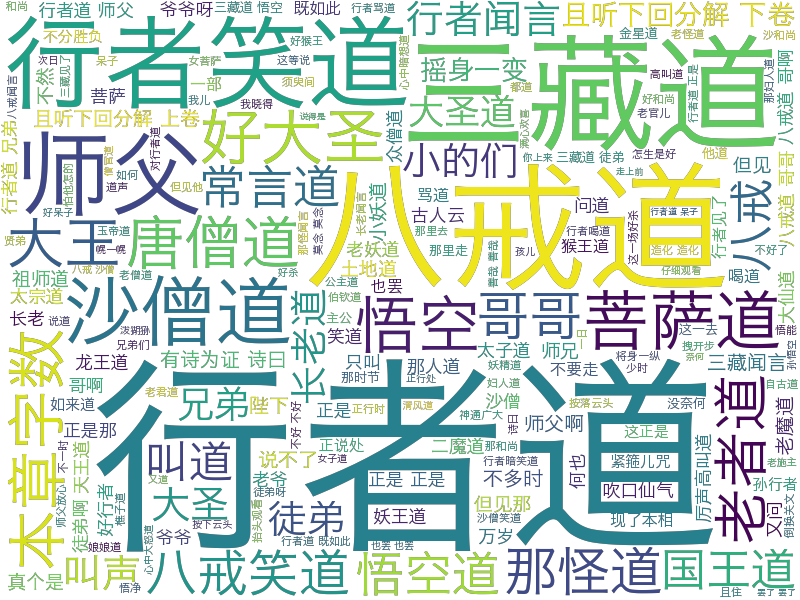
plt.imshow(wc,interpolation='bilinear')

plt.axis('off')

plt.show()

# 保存到文件

wc.to\_file(path + 'create\_images//wordcloud2.png')



**（2）中文分词效果**

这里用jieba进行中文分词

# \_\*\_ coding:utf-8 \_\*\_

# 中文分词后的词云

from wordcloud import WordCloud

import matplotlib.pyplot as plt

import jieba

path = 'D://BaiduNetdiskDownload//深度有趣人工智能//实战一 词云的使用//'

# 打开文本

text = open(path + 'source//xyj.txt',encoding = 'UTF-8').read()

# 中文分词

text = ' '.join(jieba.cut(text))

print(text[:100])

# 生成对象

wc = WordCloud(font\_path='Hiragino.ttf', width=800, height=600, mode='RGBA', background\_color=None).generate(text)

# 显示词云

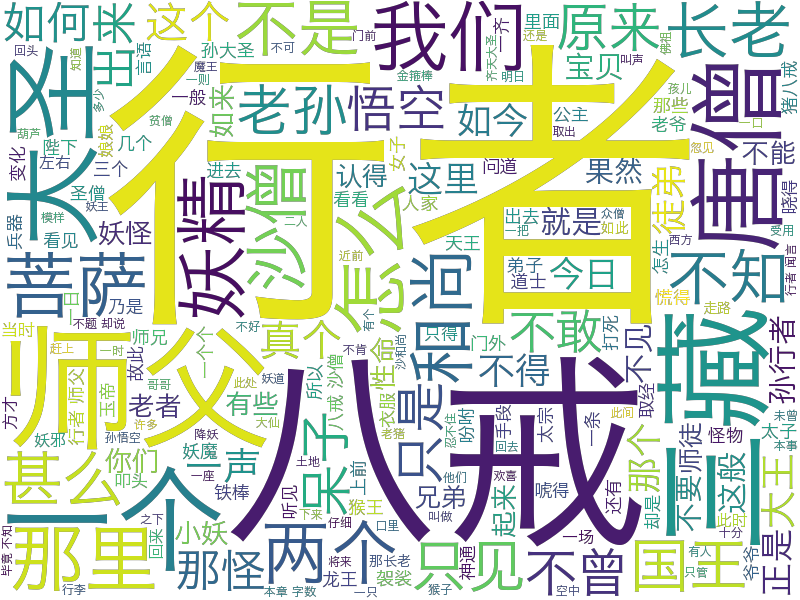
plt.imshow(wc,interpolation='bilinear')

plt.axis('off')

plt.show()

# 保存到文件

wc.to\_file(path+'create\_images//wordcloud3.png')



**（3）使用蒙版**

这里讲mask翻译为蒙版，使用蒙版之后，可以根据提供的蒙版图片，生成指定形状的的词云

# \_\*\_ coding:utf-8 \_\*\_

# 使用蒙版

from wordcloud import WordCloud

from PIL import Image

import numpy as np

import matplotlib.pyplot as plt

import jieba

path = 'D://BaiduNetdiskDownload//深度有趣人工智能//实战一 词云的使用//'

# 打开文本

text = open(path+'source//xyj.txt',encoding='utf-8').read()

# 中文分词

text = ''.join(jieba.cut(text))

print(text[:100])

# 生成对象

mask = np.array(Image.open(path+"source//black\_mask.png"))

wc = WordCloud(mask=mask,font\_path='Hiragino.ttf', width=800, height=600, mode='RGBA', background\_color=None).generate(text)

# 显示词云

plt.imshow(wc,interpolation='bilinear')

plt.axis('off')

plt.show()

# 保存到文件

wc.to\_file(path+"create\_images//wordcloud4.png")



**（4）从蒙版中抽取颜色**

# \_\*\_ coding:utf-8 \_\*\_

# 词云的颜色从蒙版中抽取

from wordcloud import WordCloud,ImageColorGenerator

from PIL import Image

import numpy as np

import matplotlib.pyplot as plt

import jieba

path = 'D://BaiduNetdiskDownload//深度有趣人工智能//实战一 词云的使用//'

# 打开文本

text = open(path+'source//xyj.txt',encoding='utf-8').read()

# 中文分词

text = ''.join(jieba.cut(text))

print(text[:100])

# 生成对象

mask = np.array(Image.open(path+"source//color\_mask.png"))

wc = WordCloud(mask=mask,font\_path='Hiragino.ttf', width=800, height=600, mode='RGBA', background\_color=None).generate(text)

# 从图片中生成颜色

image\_colors = ImageColorGenerator(mask)

wc.recolor(color\_func=image\_colors)

# 显示词云

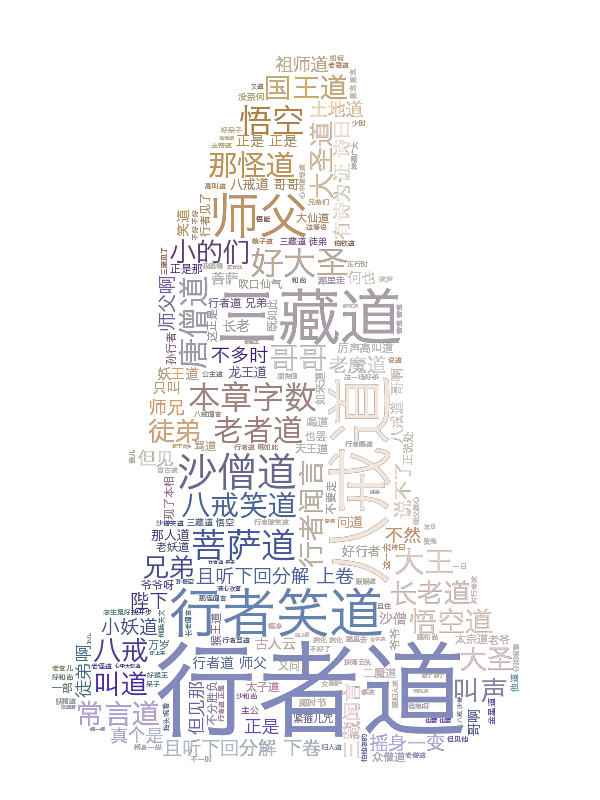
plt.imshow(wc,interpolation='bilinear')

plt.axis('off')

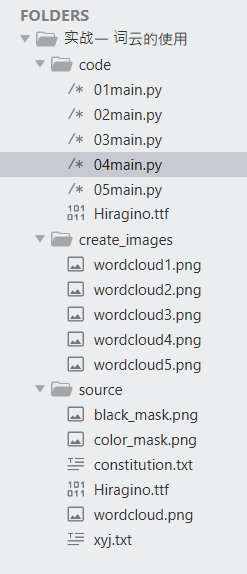
plt.show()

# 保存到文件

wc.to\_file(path+"create\_images//wordcloud5.png")



**6.实战一：手把手教你中英文词云**



项目链接：链接: https://pan.baidu.com/s/1p-QFD3WchxrvIlCnGuexKQ 提取码: zhcf 复制这段内容后打开百度网盘手机App，操作更方便哦