wordcloud库——词云绘制

1.介绍

• 词云又叫文字云,是对文本数据中出现频率较高的"关键词"在视觉上的突出呈现,形成关键词的渲染形成类似云一样的彩色图片,从而一眼就可以领略文本数据的主要表达意思。

2.安装与使用

- pip install wordcloud
- python开发环境、jieba、matplotlib、numpy、PIL 等库文件安装好。
- wordcloud.WordCloud()代表一个对应的词云对象
- 根据文本中词语出现的频率等参数绘制词云
- 绘制词云的形状、尺寸和颜色都可以设定
- 使用步骤
 - 导入模块: import wordcloud
 - 创建对象: w=wordcloud.WordCloud()
 - 。 配置参数:
 - 加载文本: w.generate("以空格分隔的文本")
 - 输出图像: w.to_file(filename)
 - 。 界面显示:

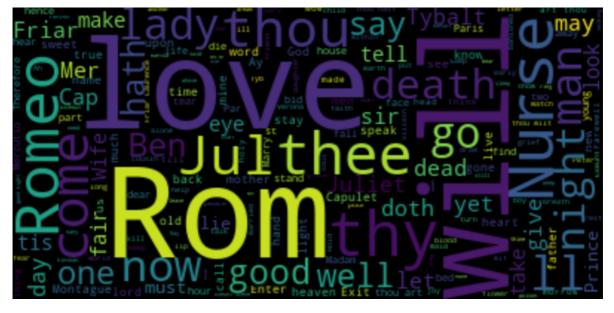
```
import matplotlib.pyplot as plt
plt.imshow(wc)
plt.axis("off")
plt.show()
```

```
import jieba
import wordcloud

# 分词

text=open("shakespare.txt",'r',encoding='utf-8').read()
ls=jieba.lcut(text)
words=" ".join(ls)

#词云
wc=wordcloud.WordCloud()
wc.generate(words)
wc.to_file("wc.png")
```



- 实际操作过程
 - 空格分割
 - 。 统计单词次数并过滤
 - 。 根据统计配置字号
 - 。 颜色环境尺寸

3.参数配置

• 对象参数配置(在创建是指定参数)

w=wordcloud.WordCloud([参数])

- o scale=10清晰度
- o width=400
- o height=200
- o min_font_size=4
- max_font_size
- o font_step=1
- o font_path="msyh.ttc"
- o max_words=200显示的最大单词数量
- stop_words={}排除次列表
- o mask词云形状
- background_color="black"词云图片背景颜色

#coding=utf-8

import jieba
import wordcloud

分词

text="奋斗创造历史,实干成就未来。我们要更加紧密地团结在以习近平同志为核心的党中央周围,高举中国特色社会主义伟大旗帜,以习近平新时代中国特色社会主义思想为指导,迎难而上,开拓进取,以经济社会发展的优异成绩迎接中华人民共和国成立70周年,为决胜全面建成小康社会、夺取新时代中国特色社会主义伟大胜利,为把我国建设成为富强民主文明和谐美丽的社会主义现代化强国、实现中华民族伟大复兴的中国梦不懈奋斗!"

ls=jieba.lcut(text)
words=" ".join(ls)

词云

wc=wordcloud.wordCloud(font_path="msyh.ttc",max_words=20,width=600,heig
ht=400,background_color="white")
wc.generate(words)
wc.to_file("wc.png")

• 词云形状mask

要求给定的图片背景是白色

```
from scipy.misc import imread
mk=imread("*.png")
wc=wordcloud.wordcloud(mask=mk)
```

或

import matplotlib.pyplot as plt
mk=plt.imread("China.jpeg")
wc=wordcloud.WordCloud(mask=mk)

#coding=utf-8

import jieba
import wordcloud
from scipy.misc import imread

分词

text="奋斗创造历史,实干成就未来。我们要更加紧密地团结在以习近平同志为核心的党中央周围,高举中国特色社会主义伟大旗帜,以习近平新时代中国特色社会主义思想为指导,迎难而上,开拓进取,以经济社会发展的优异成绩迎接中华人民共和国成立70周年,为决胜全面建成小康社会、夺取新时代中国特色社会主义伟大胜利,为把我国建设成为富强民主文明和谐美丽的社会主义现代化强国、实现中华民族伟大复兴的中国梦不懈奋斗!"

ls=jieba.lcut(text)

```
words=" ".join(ls)

# 背景
mask=imread("China.jpeg")

# 词云
wc=wordcloud.WordCloud(font_path="msyh.ttc",background_color="white",mask=mask)
wc.generate(words)
wc.to_file("wc.png")
```





注意

- 。 中文需要加上参数font_path="msyh.ttc"
- 。 界面显示时需要先imshow再show

```
#coding=utf-8
import jieba
import wordcloud
from scipy.misc import imread
import matplotlib.pyplot as plt
# 分词
text=open("gov.txt",'r',encoding='gbk').read()
ls=jieba.lcut(text)
words=" ".join(ls)
# 背景
#mask=imread("China.jpeg")
mask=plt.imread("China.jpeg")
wc=wordcloud.WordCloud(font_path="msyh.ttc",width=800,height=600,background_colo
r="white",mask=mask)
wc.generate(words)
#保存文件
wc.to_file("wc.png")
#界面显示
plt.imshow(wc)
plt.axis("off")
plt.show()
```

