Python程序设计

案例: 词云



张华 WHU

词云

■ 需求

- ☀ 一篇新闻报道,出现了哪些词?
- ₩哪些词出现得最多?
- *以"词云"的方式直观显示这些关键词。



词云

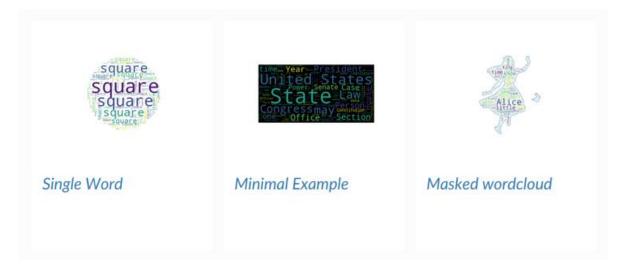


- * 文本是保存在文件中的一个长字符串
- * 从文本中分离出单词
- *用词云来显示这些词,出现次数多的词更大,出现次数少的也小

创建词云的扩展库



- *介绍和文档
 - https://amueller.github.io/word_cloud



- *****代码
 - https://github.com/amueller/word_cloud

创建词云的扩展库



- ***** 安装
 - ▶ Anaconda没有集成词云扩展库,需要自己安装。
 - ▶ 执行Anaconda Prompt
 - ➤输入命令(确保计算机可以访问Internet)
 pip install wordcloud



创建词云的扩展库

wordcloud

- #测试
 - ➤在Python解释器中输入
 >>>import wordcloud
 - >如果没有错误,则继续输入
 - >>>wc = wordcloud.WordCloud(width=320, height=240)
 - >>>wcim = wc.generate('Python computer language great high level Python easy useful Python')
 - >>>wcim.to_image().show()
 - >生成词云图像文件wc.png
 - >>>wcim.to_image().save('wc.png')
 - 图像文件在启动Python的文件夹中
 - 也可以指定保存文件的位置
 - >>>wcim.to_image().save('d:/wc.png')





创建新闻报道的词云

■ 实现思路

*步骤1:读取文件、分词整理

☀步骤2:设置并输出词云

*步骤3:观察结果,优化迭代



创建新闻报道的词云

■ 实现

```
import jieba
import wordcloud
f = open('坚持以生为本 助推学生发展.txt', 'r', encoding='utf-8')
txt = f.read()
f.close()
wordList = jieba.lcut(txt)
wcstr = ' '.join(wordList)
wc = wordcloud.WordCloud(
    r'msyh.ttc', width=640, height=480,
    background color='white',
    max words=50)
wcim = wc.generate(wcstr)
wcim.to image().save('wc news.png')
```

自定义词云

■ 查阅资料,学习使用更多的参数

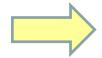
*例如,创建异形词云





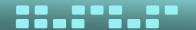
蒙板图像 (mask)

背景色为白色



异形词云

词越多,形越显



自定义词云

■ 创建异形词云

```
import jieba
import wordcloud
from imageio import imread
f = open('坚持以生为本 助推学生发展.txt', 'r', encoding='utf-8')
txt = f.read()
f.close()
wordList = jieba.lcut(txt)
wcstr = ' '.join(wordList)
maskim = imread('whulib-logo.jpg')
wc = wordcloud.WordCloud(
    'msyh.ttc', width=640, height=480,
    background color='white', mask=maskim)
wcim = wc.generate(wcstr)
wcim.to image().save('wc news whulib.png')
```