Python用于数据分析方面的应用

1. Python代码生成任意形状的文字云图

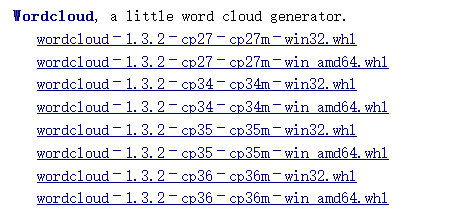
# 需要安装模块jieba和worldcloud

## jieba直接利用pip install jieba即可安装成功；

## worldcloud模块安装：

需要下载whl安装包，然后利用pip install \*\*\*.whl安装。

<http://www.lfd.uci.edu/~gohlke/pythonlibs/>



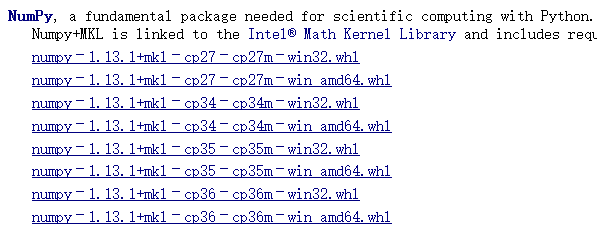
# numpy+mkl安装

下载whl包，利用pip install \*\*\*.whl安装。

<http://www.lfd.uci.edu/~gohlke/pythonlibs/#numpy>

**注意：这里利用pip install numpy安装的只是numpy，不是numpy+mkl。**

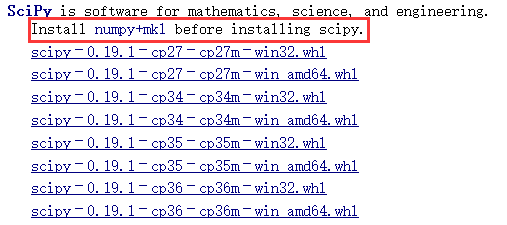
[**NumPy**](https://github.com/numpy/numpy), a fundamental package needed for scientific computing with Python.  
Numpy+MKL is linked to the [Intel® Math Kernel Library](https://software.intel.com/en-us/intel-mkl/) and includes required DLLs in the numpy.core directory.



# scipy安装

<http://www.lfd.uci.edu/~gohlke/pythonlibs/#scipy>

**scipy依赖于numpy+mkl**，所以先安装numpy+mkl。



# 错误情况及解决办法

错误： [170716 from numpy.\_distributor\_init import NUMPY\_MKL # requires numpy+mkl](http://blog.csdn.net/qq_33039859/article/details/75204962)

解决办法：

# uninstall old

**pip uninstall scipy**

**pip uninstall numpy**

# install new

**pip install numpy\_absolute\_path#下载的numpy+mkl的whl包的绝对路径**

**pip install scipy\_absolute\_path#下载的scipy的whl包的绝对路径**

## 文字云图



## 10行代码生成文字云图

# coding:utf-8  
#分词模块  
import jieba  
#画图模块  
import matplotlib.pyplot as plt  
#文字云模块  
from wordcloud import WordCloud  
#路径模块  
from os import path  
#这个是读取背景图片  
from scipy.misc import imread  
  
#找到当前路径  
d = path.dirname('.')  
#源文件路径设置  
word\_path = path.join(d,'wordcloud.txt')  
#读取文件内容  
word\_content = open(word\_path,'r',encoding='utf-8').read()  
print('##########################')  
print(word\_content)  
print('##########################')  
#设置背景图片  
mask\_img = imread(path.join(d,'heart.jpg'))  
#进行分词，好简单  
word\_cut = jieba.cut(word\_content)  
#把分词用空格连起来  
word\_cut\_join = " ".join(word\_cut)  
print('$$$$$$$$$$$$$$$$$$$')  
print(word\_cut\_join)  
print('$$$$$$$$$$$$$$$$$$$')  
wc = WordCloud(  
font\_path='simkai.ttf', #设置字体  
max\_words=50, #词云显示的最大词数  
mask=mask\_img, #设置背景图片  
background\_color='white', #背景颜色，白色好看  
).generate(word\_cut\_join)  
  
#还可以设置大小 width=2000,height=800,  
# max\_font\_size=80,min\_font\_size=16  
  
plt.imshow(wc)  
#去掉坐标轴  
plt.axis("off")  
#生成图片  
plt.show()

至于生成的思路，就比较简单了，先运用jieba分词对文章进行分词，同时读取背景的图片（用于绘制不同形状的文字图云），然后通过WordCloud生成文字图云并用画图工具进行展示。只有10行代码，精妙的文字云图油然而生。

## 乱码解决方法

刚开始的时候，文字云那一步没有设置字体，直接用**WordCloud().gengerate(word\_cut\_join)**，发现生成的是乱码，网上搜索找到解决方法，是wordcloud模块本身没有汉字字体，怎么办呢？把电脑中的默认的字体文件拷贝过来一份就可以了。

把 **C:\Windows\Fonts** 中的某种字体文件（.TTF文件）复制到当前工程路径即可，如代码中所示。