Python编程新思维及实战

嵩天

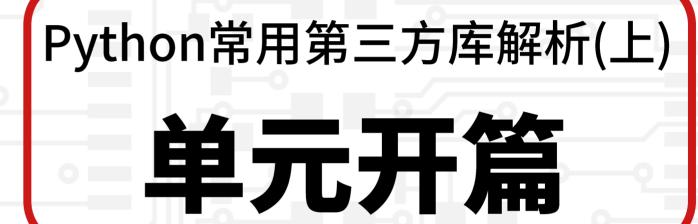


Python常用第三方库解析(上)

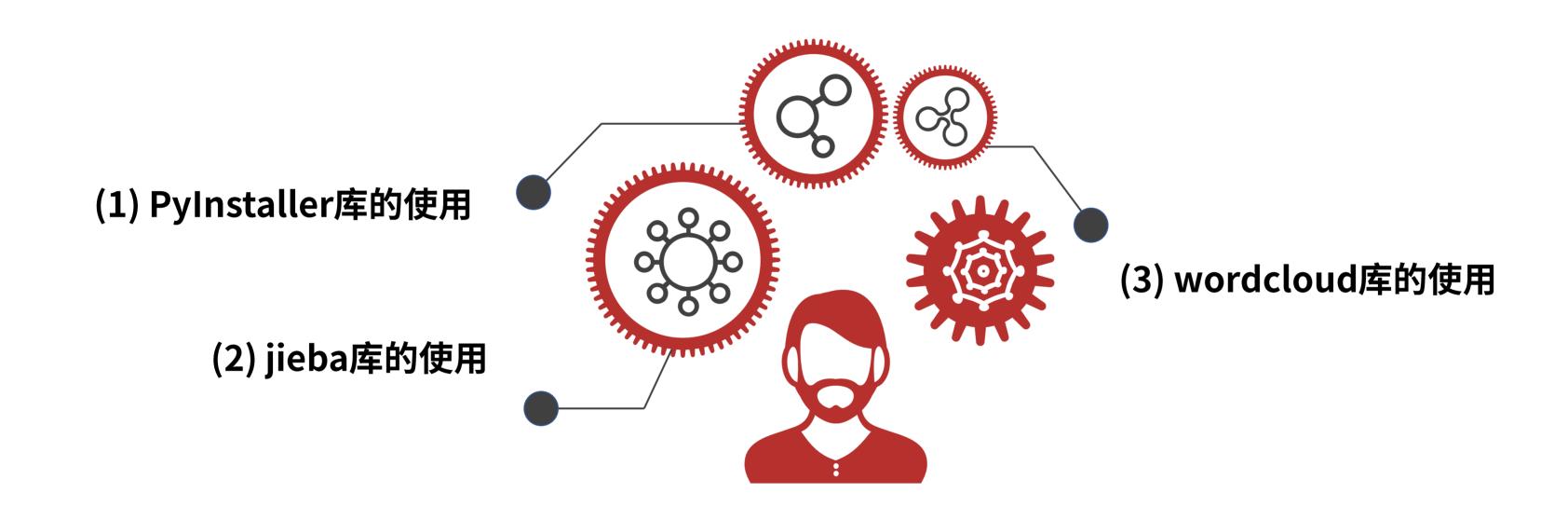
嵩天







单元开篇



Python常用第三方库解析(上)



单元开篇

通过这3个第三方库,达到如下目的

- 掌握3个常用第三方库的使用,并学会编写相关程序
- 了解第三方库使用的多样性
- 结合自身条件,了解第三方库的学习曲线



PyInstaller库介绍

PyInstaller库提供将.py源代码转换成可执行文件的功能

Mac OS X

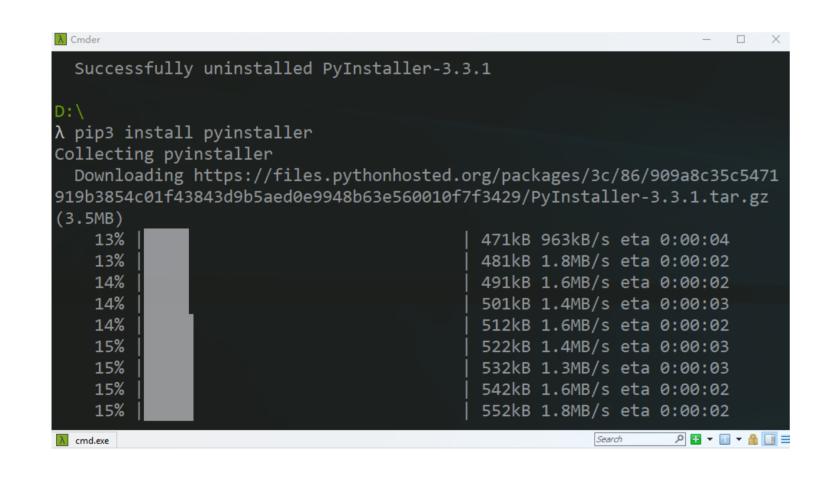
Linux

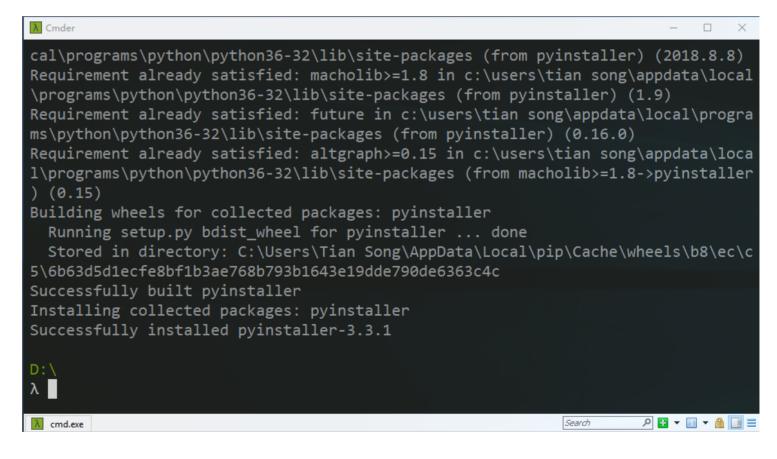
· Windows (exe文件)

https://pyinstaller.readthedocs.io

PyInstaller库安装

pip install pyinstaller









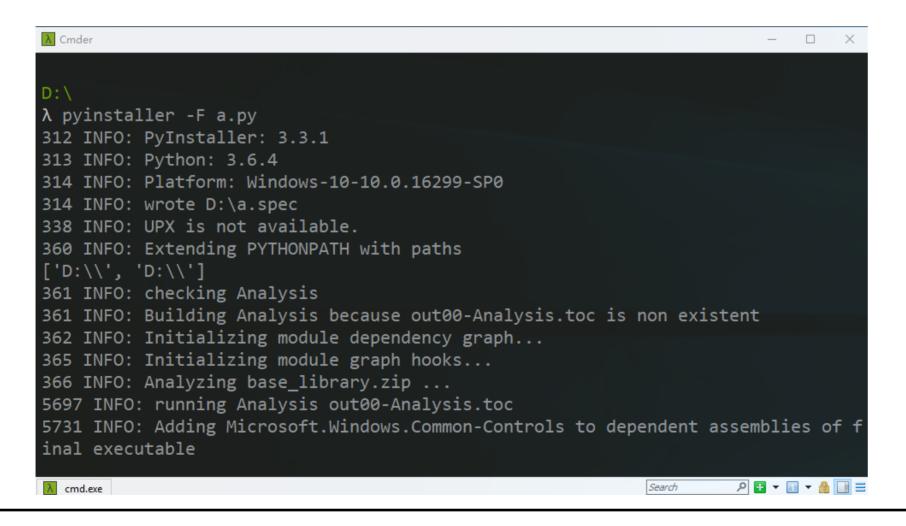
PyInstaller库介绍

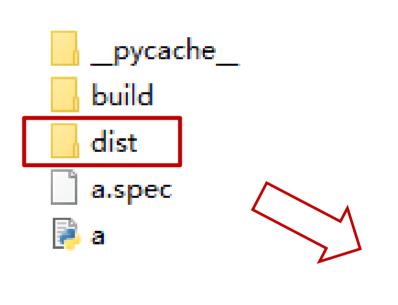
PyInstaller库提供将.py源代码转换成可执行文件的功能

- Pyinstaller是一个命令行工具,不是开发用Python模块
- Pyinstaller能够使程序脱离解释器运行,还能够防止源代码泄露
- Pyinstaller是目前最好用的Python打包工具

简单的使用

(cmd命令行) pyinstaller -F <文件名.py>







命令说明

pyinstaller [options] script [script ...]

选项(options)	选项说明
-h 或help	显示帮助信息
-v 或 -version	显示版本信息
distpath DIR	指定输出目录,默认是./dist
workpath WDIR	指定存储临时文件的目录,默认是./build
-y 或noconfirm	强制替换输出目录内容,没有确认询问

命令说明

pyinstaller [options] script [script ...]

选项(options)	选项说明
-D 或onedir	创建一个目录,包含可执行程序和程序依赖文件
-F 或onefile	创建一个独立的可执行文件
-n NAME 或 –name NAME	指定输出文件的名称目录
-i <图标文件名.ico>	指定打包程序使用的图标(icon)文件

常用指令说明

pyinstaller -F a.py

生成一个文件a.exe

pyinstaller –i b.ico -F a.py

生成一个带有图标的a.exe

pyinstaller -F a.py -n useful.exe

生成一个文件useful.exe

pyinstaller --distpath ./ -F a.py -n useful.exe 当前目录下生成useful.exe

PyInstaller库小结

pyinstaller [options] script [script ...]

- -F、-D、-h、-v、-y
- -n NAME、-i <图标文件名.ico>
- --distpath DIR、--workpath WDIR



jieba库介绍

jieba库提供中文分词功能

- 中文文本需要通过分词获得单个的词语信息
- · jieba库提供三种分词模式,支持自定义词典

https://github.com/fxsjy/jieba

jieba库安装

pip install jieba

pip3 install jieba

jieba库介绍

jieba库的三种分词模式

- 精确模式:将文本精确地切开,分词后无冗余,适合文本分词
- 全模式: 把文本中所有可能成词的词语都扫描出来, 速度快, 不能解决歧义
- 搜索引擎模式: 在精确模式基础上,对长词再次切分,提高召回率

函数	描述
jieba.cut(s)	精确模式,分词后返回一个迭代器,用for in形式遍历结果 >>>jieba.cut("新时代中国特色社会主义") <generator 0x05c90090="" at="" object="" tokenizer.cut=""></generator>
jieba.lcut(s)	精确模式,分词后返回一个列表 >>>jieba.lcut("新时代中国特色社会主义") ['新', '时代', '中国', '特色', '社会主义']

函数	描述
jieba.cut(s, cut_all=True)	全模式,分词后返回一个迭代器,用for in形式遍历结果 >>>jieba.cut("新时代中国特色社会主义", cut_all=True) <generator 0x05c90090="" at="" object="" tokenizer.cut=""></generator>
jieba.lcut(s, cut_all=True)	全模式,分词后返回一个列表 >>>jieba.lcut("新时代中国特色社会主义", cut_all=True) ['新', '时代', '中国', '国特', '特色', '社会', '社会主义', '会主', '主义']

函数	描述
jieba.cut_for_search(s)	搜索引擎模式,分词后返回一个迭代器,用for in形式遍历结果 >>>jieba.cut_for_search("新时代中国特色社会主义") <generator 0x05c90090="" at="" object="" tokenizer.cut=""></generator>
jieba.lcut_for_search(s)	搜索引擎模式,分词后返回一个列表 >>>jieba.lcut_for_search ("新时代中国特色社会主义") ['新', '时代', '中国', '特色', '社会', '会主', '主义', '社会主义']

函数	描述
jieba.add_word(w)	向分词词典增加新词w >>>jieba.add_word("蟒蛇语言") >>>jieba.lcut("Python也被称作蟒蛇语言") ['Python', '也', '被称作', '蟒蛇语言']

jieba库小结

jieba库的使用

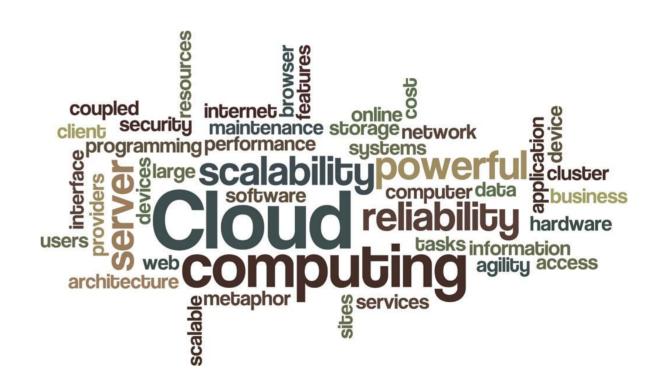
- 精确模式: cut(s)、lcut(s)
- 全模式: cut(s, cut_all = True)、lcut(s, cut_all = True)
- 搜索引擎模式: cut_for_search(s)、lcut_for_search(s)
- 动态增加新词: add_word(w)



wordcloud库介绍

wordcloud库提供词云展示功能





https://github.com/amueller/word_cloud

wordcloud库安装

pip install wordcloud

部分Windows版本: UCI安装方法

wordcloud库介绍

wordcloud库把词云当作一个WordCloud对象

- wordcloud.WordCloud()代表一个文本对应的词云
- 可以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、尺寸和颜色等可以设定

- 以WordCloud对象为基础
- 配置参数、加载文本、输出文件

方法	描述
w.generate(txt)	向WordCloud对象w中加载文本txt >>>w.generate("One Python, One World")
w.to_file(filename)	将词云输出为图像文件,.png或.jpg格式 >>>w.to_file("outfile.png")

wordcloud库的使用步骤

```
import wordcloud
```

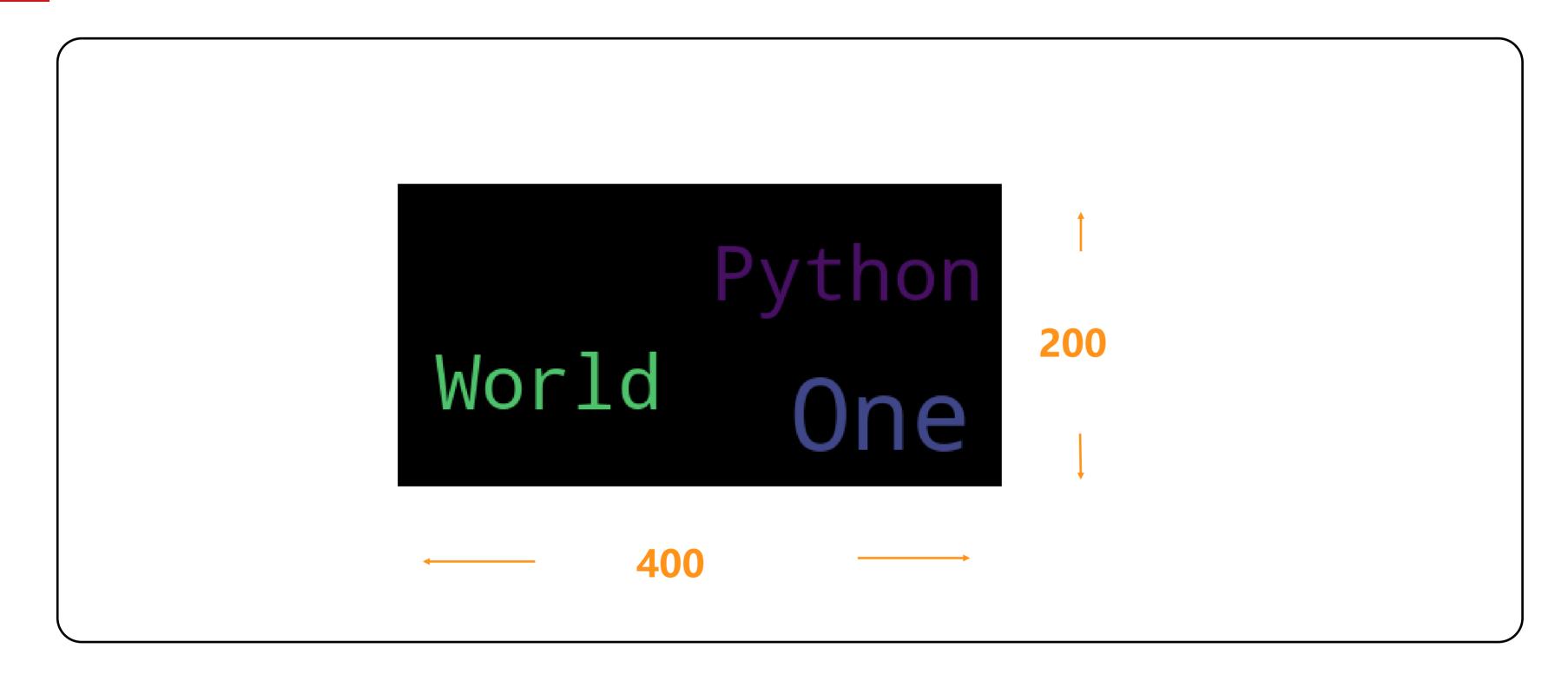
```
c = wordcloud.WordCloud()
```

• 步骤1: 配置对象参数

```
c.generate("One Python, One World") · 步骤2: 加载词云文本
```

```
c.to file("outfile.png")
```

• 步骤3: 输出词云文件



wordcloud库基本流程

"One Python, One World"



Python World One

文本



① 分隔: 以空格分隔单词

② 统计: 单词出现次数并过滤

③ 字体: 根据统计配置字号

④ 布局: 颜色环境尺寸



词云



参数	描述
width	指定词云对象生成图片的宽度,默认400像素 >>>w=wordcloud.WordCloud(width=600)
height	指定词云对象生成图片的高度,默认200像素 >>>w=wordcloud.WordCloud(width=400)

参数	描述
min_font_size	指定词云中字体的最小字号,默认4号 >>>w=wordcloud.WordCloud(min_font_size=10)
max_font_size	指定词云中字体的最大字号,根据高度自动调节 >>>w=wordcloud.WordCloud(min_font_size=20)



参数	描述
font_step	指定词云中字体字号的步进间隔,默认为1 >>>w=wordcloud.WordCloud(font_step=2)
font_path	指定字体文件的路径,默认None >>>w=wordcloud.WordCloud(font_path="msyh.ttc")

参数	描述
max_words	指定词云显示的最大单词数量,默认200 >>>w=wordcloud.WordCloud(max_words=20)
stop_words	指定词云的排除词列表,即不显示的单词列表 >>>w=wordcloud.WordCloud(stop_words={"Python"})

参数	描述
mask	指定词云形状,需要引用imread()函数,默认为长方形 >>>from scipy.misc import imread >>>mk=imread("pic.png") #参考图片的背景是白色 >>>w=wordcloud.WordCloud(mask=mk)
background_color	指定词云图片的背景颜色,默认为黑色 >>>w=wordcloud.WordCloud(background_color="white")

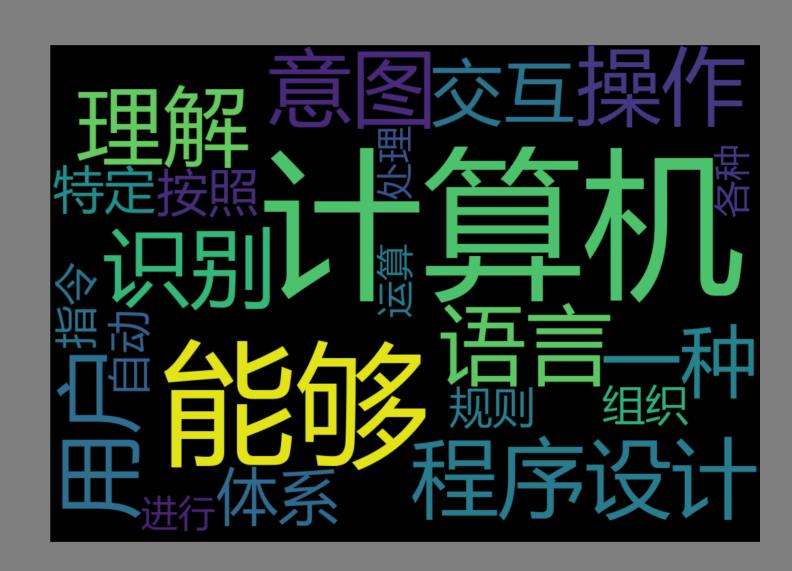
wordcloud库实例

```
import wordcloud
txt = "life is short, you need python"
w = wordcloud. WordCloud( \
       background color = "white")
w. generate(txt)
w. to file("pywcloud.png")
```



wordcloud库实例

```
import jieba
import wordcloud
txt = "程序设计语言是计算机能够理解和\
识别用户操作意图的一种交互体系,它按照\
特定规则组织计算机指令,使计算机能够自\
动进行各种运算处理。"
w = wordcloud. WordCloud( width=1000, \
   font_path="msyh. ttc", height=700)
w. generate(" ". join(jieba. lcut(txt)))
w. to_file("pywcloud.png")
```



中文需要先分词并组成空格分隔字符串





wordcloud库小结

w = wordcloud.WordCloud(<参数>)

- 两个方法: w.generate(txt)、w.to_file(filename)
- 参数列表: width、height、min_font_size、 max_font_size

font_step、font_path、max_words、stop_words

mask, background_color

Python常用第三方库解析(上)

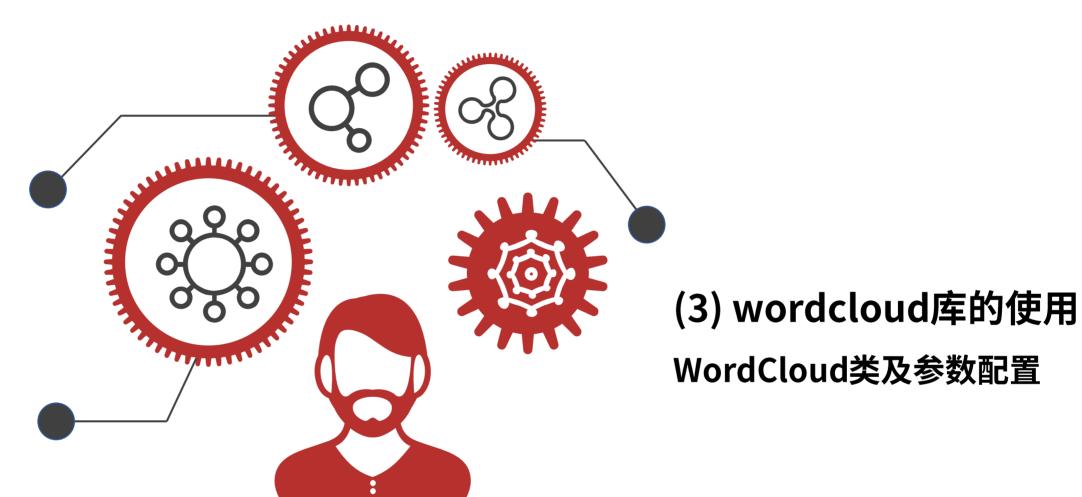
单元小结



单元开篇

(1) PyInstaller库的使用 Pyinstaller命令参数: -F、-i等

(2) jieba库的使用 精确模式、全模式、搜索引擎模式



Python常用第三方库解析(上)



PyInstaller库小结

pyinstaller [options] script [script ...]

- -F、-D、-h、-v、-y
- -n NAME、-i <图标文件名.ico>
- --distpath DIR、--workpath WDIR

jieba库小结

jieba库的使用

- 精确模式: cut(s)、lcut(s)
- 全模式: cut(s, cut_all = True)、lcut(s, cut_all = True)
- 搜索引擎模式: cut_for_search(s)、lcut_for_search(s)
- 动态增加新词: add_word(w)

wordcloud库小结

w = wordcloud.WordCloud(<参数>)

- 两个方法: w.generate(txt)、w.to_file(filename)
- 参数列表: width、height、min_font_size、 max_font_size

font_step、font_path、max_words、stop_words

mask, background_color



Thank you