

python

The Python logo, consisting of two interlocking snakes, one blue and one yellow, is positioned below the word "python".

```
import turtle
turtle.setup(650,350,200,200)
turtle.penup()
turtle.fd(-250)
turtle.pendown()
turtle.pensize(25)
turtle.pencolor("purple")

for i in range(4):
    turtle.circle(40, 80)
    turtle.circle(-40, 80)
    turtle.circle(40, 80/2)
    turtle.fd(40)
    turtle.circle(16, 180)
    turtle.fd(40 * 2/3)
```

Python语言程序设计

模块6: wordcloud库的使用



嵩 天
北京理工大学





wordcloud库基本介绍

wordcloud库概述

wordcloud是优秀的词云展示第三方库



- 词云以词语为基本单位，更加直观和艺术的展示文本

wordcloud库的安装

(cmd命令行) `pip install wordcloud`

```
命令提示符
Microsoft Windows [版本 10.0.16299.371]
(c) 2017 Microsoft Corporation. 保留所有权利。

C:\Users\Tian Song>pip install wordcloud
Collecting wordcloud
  Downloading https://files.pythonhosted.org/packages/bc/e8/cab8479b25297b3347cfb55e85a5014e8c53b80e513eaf1ba58c7b3a6acd/wordcloud-1.4.1.tar.gz (172kB)
    100% |#####| 174kB 33kB/s
Installing collected packages: wordcloud
Successfully installed wordcloud-1.4.1
```

```
命令提示符
python36-32\lib\site-packages (from wordcloud)
Requirement already satisfied: numpy>=1.6.1 in c:\users\tian song\appdata\local\programs\python\python36-32\lib\site-packages (from wordcloud)
Requirement already satisfied: pillow in c:\users\tian song\appdata\local\programs\python\python36-32\lib\site-packages (from wordcloud)
Requirement already satisfied: six>=1.10 in c:\users\tian song\appdata\local\programs\python\python36-32\lib\site-packages (from matplotlib->wordcloud)
Requirement already satisfied: pytz in c:\users\tian song\appdata\local\programs\python\python36-32\lib\site-packages (from matplotlib->wordcloud)
Requirement already satisfied: cycler>=0.10 in c:\users\tian song\appdata\local\programs\python\python36-32\lib\site-packages (from matplotlib->wordcloud)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\tian song\appdata\local\programs\python\python36-32\lib\site-packages (from matplotlib->wordcloud)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in c:\users\tian song\appdata\local\programs\python\python36-32\lib\site-packages (from matplotlib->wordcloud)
Requirement already satisfied: olefile in c:\users\tian song\appdata\local\programs\python\python36-32\lib\site-packages (from pillow->wordcloud)
Installing collected packages: wordcloud
  Running setup.py install for wordcloud ... done
Successfully installed wordcloud-1.4.1
You are using pip version 9.0.1, however version 10.0.0 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.

C:\Users\Tian Song>
```



wordcloud库使用说明



wordcloud库基本使用

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、尺寸和颜色都可以设定**

wordcloud库常规方法

w = wordcloud.WordCloud()

- 以WordCloud对象为基础
- 配置参数、加载文本、输出文件

wordcloud库常规方法

w = wordcloud.WordCloud()

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

wordcloud库常规方法

```
import wordcloud
```

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

```
c.generate("wordcloud by Python")
```

```
c.to_file("pywordcloud.png")
```

- 步骤1：配置对象参数

- 步骤2：加载词云文本

- 步骤3：输出词云文件

wordcloud库常规方法



↑
200
↓

← 400 →

wordcloud库常规方法

"wordcloud by Python"



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

配置对象参数

w = wordcloud.WordCloud(<参数>)

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)

配置对象参数

参数	描述
min_font_size	指定词云中字体的最小字号，默认4号 <code>>>>w=wordcloud.WordCloud(min_font_size=10)</code>
max_font_size	指定词云中字体的最大字号，根据高度自动调节 <code>>>>w=wordcloud.WordCloud(min_font_size=20)</code>
font_step	指定词云中字体字号的步进间隔，默认为1 <code>>>>w=wordcloud.WordCloud(font_step=2)</code>

配置对象参数

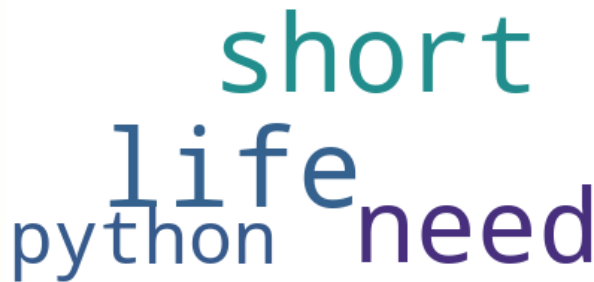
参数	描述
font_path	指定字体文件的路径，默认None <code>>>>w=wordcloud.WordCloud(font_path="msyh.ttc")</code>
max_words	指定词云显示的最大单词数量，默认200 <code>>>>w=wordcloud.WordCloud(max_words=20)</code>
stop_words	指定词云的排除词列表，即不显示的单词列表 <code>>>>w=wordcloud.WordCloud(stop_words={"Python"})</code>

配置对象参数

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

wordcloud应用实例

```
import wordcloud  
txt = "life is short, you need python"  
w = wordcloud.WordCloud( \  
    background_color = "white")  
w.generate(txt)  
w.to_file("pywcloud.png")
```



short
life
python need

以空格分隔单词

```
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



