

# Python语言总结

版本：v2.0

Crifan Li

## 摘要

本文主要介绍了个人目前已经发布整理的Python资料，以及和Python语言本身相关的总结，以及各种内置（str,unicode,json,OptionParser等）模块的总结，和第三方（Beautifulsoup等）库的总结，以及介绍Python方面的有价值的参考资料。



## 本文提供多种格式供：

在线阅读	<a href="#">HTML</a> <sup>1</sup>	<a href="#">HTMLs</a> <sup>2</sup>	<a href="#">PDF</a> <sup>3</sup>	<a href="#">CHM</a> <sup>4</sup>	<a href="#">TXT</a> <sup>5</sup>	<a href="#">RTF</a> <sup>6</sup>	<a href="#">WEBHELP</a> <sup>7</sup>
下载（7zip压缩包）	<a href="#">HTML</a> <sup>8</sup>	<a href="#">HTMLs</a> <sup>9</sup>	<a href="#">PDF</a> <sup>10</sup>	<a href="#">CHM</a> <sup>11</sup>	<a href="#">TXT</a> <sup>12</sup>	<a href="#">RTF</a> <sup>13</sup>	<a href="#">WEBHELP</a> <sup>14</sup>

HTML版本的在线地址为：

[http://www.crifan.com/files/doc/docbook/python\\_summary/release/html/python\\_summary.html](http://www.crifan.com/files/doc/docbook/python_summary/release/html/python_summary.html)

有任何意见，建议，提交bug等，都欢迎去讨论组发帖讨论：

[http://www.crifan.com/bbs/categories/python\\_summary/](http://www.crifan.com/bbs/categories/python_summary/)

## 修订历史

修订 2.0	2013-10-10	crl
<ol style="list-style-type: none"><li>把crifanLib.py独立出去称为单独的book</li><li>把如何使用和运行Python脚本移至：python初级教程：入门详解</li><li>重新整理此book的架构</li></ol>		
修订 2012-11-06	1.2	crl
<ol style="list-style-type: none"><li>将Python从language_summary移至此独立book</li><li>添加了Python代码的格式化</li><li>添加了BeautifulSoup中find参数可以为正则表达式</li></ol>		

<sup>1</sup> [http://www.crifan.com/files/doc/docbook/python\\_summary/release/html/python\\_summary.html](http://www.crifan.com/files/doc/docbook/python_summary/release/html/python_summary.html)

<sup>2</sup> [http://www.crifan.com/files/doc/docbook/python\\_summary/release/htmls/index.html](http://www.crifan.com/files/doc/docbook/python_summary/release/htmls/index.html)

<sup>3</sup> [http://www.crifan.com/files/doc/docbook/python\\_summary/release/pdf/python\\_summary.pdf](http://www.crifan.com/files/doc/docbook/python_summary/release/pdf/python_summary.pdf)

<sup>4</sup> [http://www.crifan.com/files/doc/docbook/python\\_summary/release/chm/python\\_summary.chm](http://www.crifan.com/files/doc/docbook/python_summary/release/chm/python_summary.chm)

<sup>5</sup> [http://www.crifan.com/files/doc/docbook/python\\_summary/release/txt/python\\_summary.txt](http://www.crifan.com/files/doc/docbook/python_summary/release/txt/python_summary.txt)

<sup>6</sup> [http://www.crifan.com/files/doc/docbook/python\\_summary/release/rtf/python\\_summary.rtf](http://www.crifan.com/files/doc/docbook/python_summary/release/rtf/python_summary.rtf)

<sup>7</sup> [http://www.crifan.com/files/doc/docbook/python\\_summary/release/webhelp/index.html](http://www.crifan.com/files/doc/docbook/python_summary/release/webhelp/index.html)

<sup>8</sup> [http://www.crifan.com/files/doc/docbook/python\\_summary/release/html/python\\_summary.html.7z](http://www.crifan.com/files/doc/docbook/python_summary/release/html/python_summary.html.7z)

<sup>9</sup> [http://www.crifan.com/files/doc/docbook/python\\_summary/release/htmls/index.html.7z](http://www.crifan.com/files/doc/docbook/python_summary/release/htmls/index.html.7z)

<sup>10</sup> [http://www.crifan.com/files/doc/docbook/python\\_summary/release/pdf/python\\_summary.pdf.7z](http://www.crifan.com/files/doc/docbook/python_summary/release/pdf/python_summary.pdf.7z)

<sup>11</sup> [http://www.crifan.com/files/doc/docbook/python\\_summary/release/chm/python\\_summary.chm.7z](http://www.crifan.com/files/doc/docbook/python_summary/release/chm/python_summary.chm.7z)

<sup>12</sup> [http://www.crifan.com/files/doc/docbook/python\\_summary/release/txt/python\\_summary.txt.7z](http://www.crifan.com/files/doc/docbook/python_summary/release/txt/python_summary.txt.7z)

<sup>13</sup> [http://www.crifan.com/files/doc/docbook/python\\_summary/release/rtf/python\\_summary.rtf.7z](http://www.crifan.com/files/doc/docbook/python_summary/release/rtf/python_summary.rtf.7z)

<sup>14</sup> [http://www.crifan.com/files/doc/docbook/python\\_summary/release/webhelp/python\\_summary.webhelp.7z](http://www.crifan.com/files/doc/docbook/python_summary/release/webhelp/python_summary.webhelp.7z)

- 
4. 添加了在Mac下运行Python脚本
  5. 更新crifanLib.py的最新链接
-

---

# Python语言总结:

Crifan Li

版本 : v2.0

出版日期 2013-10-10

版权 © 2013 Crifan, <http://crifan.com>

本文章遵从 : [署名-非商业性使用 2.5 中国大陆\(CC BY-NC 2.5\)](http://creativecommons.org/licenses/by-nc/2.5/)<sup>15</sup>

---

<sup>15</sup> [http://www.crifan.com/files/doc/docbook/soft\\_dev\\_basic/release/html/soft\\_dev\\_basic.html#cc\\_by\\_nc](http://www.crifan.com/files/doc/docbook/soft_dev_basic/release/html/soft_dev_basic.html#cc_by_nc)

---

---

# 目录

1. Python简介 .....	1
1.1. Python的特点 .....	1
2. 已发布的Python的相关资料 .....	2
2.1. python初级教程：入门详解 .....	2
2.2. python中级教程：开发总结 .....	3
2.3. Python高级教程 .....	3
3. Python语言本身的理解和心得 .....	4
3.1. 学习python的本质就是学习对应的各种模块的用法 .....	4
3.2. Python代码的格式化 .....	4
3.3. 将py文件编译成pyc .....	4
3.4. for循环中的变量是只读的，不可修改 .....	5
4. python内置模块学习心得 .....	6
4.1. Python中re模块 .....	6
4.2. Python中的json .....	6
4.3. Python参数处理模块：OptionParser和argparse .....	6
4.3.1. 利用OptionParser库中的add_option添加脚本参数和帮助信息 .....	6
4.3.2. 官方建议放弃OptionParser，使用新的模块：argparse .....	7
4.4. Python中和字符串相关的内容（字符编码，str，Unicode等）的心得 .....	7
5. Python第三方模块学习心得 .....	8
6. Python相关资料 .....	9
参考书目 .....	10

---

# 第 1 章 Python简介

## 1.1. Python的特点

1. 不需要编译，即可运行
2. 有很多特性，使得语言本身很方便编程者使用

---

## 第 2 章 已发布的Python的相关资料

之前自己折腾了很多Python方面的东西。

然后，后来就整理出了很多东西：

- 系列教程  
将相关的内容，整理成系列的教程：
  1. Python初级教程  
详见：[第 2.1 节 “python初级教程：入门详解”](#)
  2. Python中级教程  
详见：[第 2.2 节 “python中级教程：开发总结”](#)
  3. Python高级教程  
详见：[第 2.3 节 “Python高级教程”](#)

- 库函数：[crifanLib.py](#)  
将一些常用的功能，整理成通用的函数。

最终形成了自己的python的库函数

详见：

[详解crifan的Python库：crifanLib.py](#)<sup>1</sup>

### 2.1. python初级教程：入门详解

关于Python方面的基础知识已移至：

[python初级教程：入门详解](#)<sup>2</sup>

包括下载合适版本的Python，安装Python等等内容。



#### 如何使用和运行Python脚本

之前的，如何使用和运行Python脚本，也已移至：

[python初级教程：入门详解](#)<sup>3</sup>

中的对应位置：

[如何在Windows环境下使用Python脚本](#)<sup>4</sup>

和：

[如何在Mac环境下运行Python脚本](#)<sup>5</sup>

<sup>1</sup> [http://www.crifan.com/files/doc/docbook/crifanlib\\_python/release/htmls/](http://www.crifan.com/files/doc/docbook/crifanlib_python/release/htmls/)

<sup>2</sup> [http://www.crifan.com/files/doc/docbook/python\\_beginner\\_tutorial/release/htmls/](http://www.crifan.com/files/doc/docbook/python_beginner_tutorial/release/htmls/)

<sup>3</sup> [http://www.crifan.com/files/doc/docbook/python\\_beginner\\_tutorial/release/htmls/](http://www.crifan.com/files/doc/docbook/python_beginner_tutorial/release/htmls/)

<sup>4</sup> [http://www.crifan.com/files/doc/docbook/python\\_beginner\\_tutorial/release/htmls/python\\_dev\\_env.html#use\\_python\\_under\\_windows](http://www.crifan.com/files/doc/docbook/python_beginner_tutorial/release/htmls/python_dev_env.html#use_python_under_windows)

<sup>5</sup> [http://www.crifan.com/files/doc/docbook/python\\_beginner\\_tutorial/release/htmls/run\\_python\\_under\\_mac.html](http://www.crifan.com/files/doc/docbook/python_beginner_tutorial/release/htmls/run_python_under_mac.html)

## 2.2. python中级教程：开发总结

后来，又整理了出了，开发过程中很多的心得，

然后把所有内容都放进了对应的教程：

[python中级教程：开发总结](#)<sup>6</sup>

## 2.3. Python高级教程

接着，对于Python，相对的一些知识点，专门整理出了一系列的教程，算是高级教程系列。

目前已有：

- [Python专题教程：字符串和字符编码](#)<sup>7</sup>
- [Python专题教程：抓取网站，模拟登陆，抓取动态网页](#)<sup>8</sup>
- [Python专题教程：正则表达式re模块详解](#)<sup>9</sup>
- [Python专题教程：BeautifulSoup详解](#)<sup>10</sup>

---

<sup>6</sup> [http://www.crifan.com/files/doc/docbook/python\\_intermediate\\_tutorial/release/htmls/](http://www.crifan.com/files/doc/docbook/python_intermediate_tutorial/release/htmls/)

<sup>7</sup> [http://www.crifan.com/files/doc/docbook/python\\_topic\\_str\\_encoding/release/htmls/](http://www.crifan.com/files/doc/docbook/python_topic_str_encoding/release/htmls/)

<sup>8</sup> [http://www.crifan.com/files/doc/docbook/python\\_topic\\_web\\_scrape/release/htmls/](http://www.crifan.com/files/doc/docbook/python_topic_web_scrape/release/htmls/)

<sup>9</sup> [http://www.crifan.com/files/doc/docbook/python\\_topic\\_re/release/htmls/](http://www.crifan.com/files/doc/docbook/python_topic_re/release/htmls/)

<sup>10</sup> [http://www.crifan.com/files/doc/docbook/python\\_topic\\_beautifulsoup/release/htmls/](http://www.crifan.com/files/doc/docbook/python_topic_beautifulsoup/release/htmls/)

---

# 第 3 章 Python语言本身的理解和心得

此处整理一下，关于Python语言本身，这方面的心得和体会：

## 3.1. 学习python的本质就是学习对应的各种模块的用法

python中有很多其他的库，帮你实现各种需要的功能，而你要做的事情，就是import对应的库，然后调用对应的函数即可。

而利用python去编程，去实现一定的功能，更多的层面上，除了学习完python的基本语法之后，就是对各种库，各种模块的如何使用上了，可以极大地提高你做事情的效率。

## 3.2. Python代码的格式化

python中，没有（类似于其他语言可用的，比如SourceFormatX之类的）代码格式化工具，即意味着，你无法用工具，打开python文件，直接点击某个按钮，就可以帮你全部格式化好了。

只能通过手工去格式化你的python代码

所谓手工，那就是，把属于每个函数（def关键字）所属代码，按照自己对代码的内在逻辑的理解，去一行行的缩进

此工作，目前只能手工做，好像也没有其他更加方便的方法。

单独针对手工一行行的处理python代码，不同的工具中，实现的效率也会不同。

推荐使用Notepad++，因为此软件有很多方便你格式化Python代码的功能：

- 可以显示所有特殊字符，包括空格，TAB键，行尾符等  
具体参考：[Notepad++支持显示回车符，换行符，TAB键，行首，行尾等特殊字符<sup>1</sup>](#)
- 支持将TAB键自动转换为空格  
目的在于，对于Python这样的语言，是靠缩进来决定代码逻辑的，所以对于缩进，TAB键和（4个连续的）空格，两者之间不是等价的，会产生语法错误的。

此时就可以利用Notepad++的，将TAB键自动转为（默认为4个，可以自定义个数）空格，实现很方便的代码格式化。

具体参考：[Notepad++支持用空格取代TAB键<sup>2</sup>](#)

总之，关于Python代码的格式化，需要手动处理Python代码，而用Notepad++去处理，可以极大地提高效率。

## 3.3. 将py文件编译成pyc

参考[这里<sup>3</sup>](#)，启动Python的IDE – IDLE (Python GUI)

然后在里面输入：

---

<sup>1</sup> [http://www.crifan.com/files/doc/docbook/crifan\\_rec\\_soft/release/html/crifan\\_rec\\_soft.html#npp\\_func\\_show\\_special\\_char](http://www.crifan.com/files/doc/docbook/crifan_rec_soft/release/html/crifan_rec_soft.html#npp_func_show_special_char)

<sup>2</sup> [http://www.crifan.com/files/doc/docbook/crifan\\_rec\\_soft/release/html/crifan\\_rec\\_soft.html#npp\\_func\\_space\\_replace\\_tab](http://www.crifan.com/files/doc/docbook/crifan_rec_soft/release/html/crifan_rec_soft.html#npp_func_space_replace_tab)

<sup>3</sup> <http://hi.baidu.com/%C1%AC%BF%B419%BC%AF/blog/item/2e3197dd8c209be476c63825.html>



```
import py_compile
```

回车 (↵) 后再输入：

```
py_compile.compile(r"E:\dev_root\Python25\Lib\sgmlib.py")
```

就可以将对应的py文件编译成pyc了，生成的sgmlib.pyc在同目录下。

### 3.4. for循环中的变量是只读的，不可修改

在使用for循环时，要注意其所得的单个变量，是只读的，不能修改。

想要修改其值的话，只能借助其他变量实现。

示例代码如下：

```
for singleContent in soupContents: #singleContent is BeautifulSoup.Tag
    .....
    if(recursive):
        .....
        filteredSingleContent = singleContent; # here must use another tmp value to store
changed value
        filteredSubContentList =
removeSoupContentsTagAttr(filteredSingleContent.contents, .....);
        .....
        filteredSingleContent.contents = filteredSubContentList; # here if use
singleContent.contents = filteredSubContentList will cause error !!!
        #logging.debug("[%d] after filter, sub contents=%s", currentLevel, filteredSingleContent);
        filteredContents.append(filteredSingleContent);
    else:
        .....
```

---

# 第 4 章 python内置模块学习心得

此处整理一些，Python内置的各种模块的使用心得：

## 4.1. Python中re模块

Python中的re模块的相关使用心得，请参见

[Python专题教程：正则表达式re模块详解](#)<sup>1</sup>

## 4.2. Python中的json

相关内容已移至：

[Python中处理JSON字符串](#)<sup>2</sup>

## 4.3. Python参数处理模块：OptionParser和argparse

### 4.3.1. 利用OptionParser库中的add\_option添加脚本参数和帮助信息

在写脚本的时候，可以利用add\_option去添加对应的参数解析以及帮助信息，而剩下的事情，如何去解析对应的参数和显示帮助信息，都是由OptionParser自动完成了。

参考如下内容：

```
from optparse import OptionParser

def main():
    #main procedure begin
    parser = OptionParser()
    parser.add_option("-s", "--source", action="store", type="string", dest="srcURL", help="source msn/live space address")
    parser.add_option("-f", "--startfrom", action="store", type="string", dest="startfromURL", help="a permalink in source msn/live space address for starting with, if this is specified, srcURL will be ignored.")
    parser.add_option("-x", "--proxy", action="store", type="string", dest="proxy", help="http proxy server, only for connecting live space.I don't know how to add proxy for metaWeblog yet. So this option is probably not useful...")

    (options, args) = parser.parse_args()
    #export all options variables
    for i in dir(options):
        exec i+" = options."+i
    #add proxy
    if proxy:
        XXX
```

<sup>1</sup> [http://www.crifan.com/files/doc/docbook/python\\_topic\\_re/release/html/python\\_topic\\_re.html](http://www.crifan.com/files/doc/docbook/python_topic_re/release/html/python_topic_re.html)

<sup>2</sup> [http://www.crifan.com/files/doc/docbook/json\\_tutorial/release/html/json\\_tutorial.html#python\\_json](http://www.crifan.com/files/doc/docbook/json_tutorial/release/html/json_tutorial.html#python_json)

```
if startfromURL :
    XXX
elif srcURL:
    XXX
else:
    logging.error("错误XXX")
    sys.exit(2)
```

然后我们就可以在运行脚本的时候，添加参数了，比如：

```
BlogsToWordpress.py -s http://hi.baidu.com/recommend_music
```

想要查看帮助信息，就是常见的-h或-help：

```
BlogsToWordpress.py -h
```

更多例子和解释，可以参考这里：

[python模块学习——optparse](#) <sup>3</sup>

### 4.3.2. 官方建议放弃OptionParser，使用新的模块： argparse

根据官网的解释，其实optparse已废弃，建议使用最新的更好用的argparse

此处之所以还继续介绍optparse主要是由于之前还是一直用的是optparse，而argparse自己只是试用过，不是很熟悉。

在可以看懂官网的例子的前提下，还是建议使用argparse

## 4.4. Python中和字符串相关的内容（字符编码，str，Unicode等）的心得

字符编码相关的内容，已整理至：

[Python专题教程：字符串和字符编码](#) <sup>4</sup>

---

<sup>3</sup> [http://www.cnblogs.com/captain\\_jack/archive/2011/01/11/1933366.html](http://www.cnblogs.com/captain_jack/archive/2011/01/11/1933366.html)

<sup>4</sup> [http://www.crifan.com/files/doc/docbook/python\\_topic\\_str\\_encoding/release/htmls/](http://www.crifan.com/files/doc/docbook/python_topic_str_encoding/release/htmls/)

---

# 第 5 章 Python第三方模块学习心得

此处整理Python的第三方的模块，库等的学习心得：

前面第 3.1 节 “[学习python的本质就是学习对应的各种模块的用法](#)” 已经说了，

其实学习Python的过程，很多时候就是在学习如何使用第三方模块，完成自己需要的功能。

下面就来总结一下，关于，Python的第三方的模块方面的学习心得和体会：

## 5.1. 去哪找python的第三方类库

关于Python的第三方库类库，其实网上很多很多相关资料。

其中，官网的Python库：[Python Package Index](#)<sup>1</sup>，其中有N多N多的库，有需要的人，可以去那里找找。

其他的网上的N多资源中，我觉得值得看看的有：

1. <http://www.elias.cn/Python/HomePage>中的3.3 用第三方类库的部分。

## 5.2. BeautifulSoup

关于BeautifulSoup相关的内容已整理至独立教程：

[Python专题教程：BeautifulSoup详解](#)<sup>2</sup>

---

<sup>1</sup> <http://pypi.python.org/pypi/>

<sup>2</sup> [http://www.crifan.com/files/doc/docbook/python\\_topic\\_beautifulsoup/release/html/python\\_topic\\_beautifulsoup.html](http://www.crifan.com/files/doc/docbook/python_topic_beautifulsoup/release/html/python_topic_beautifulsoup.html)

---

# 第 6 章 Python相关资料

此处记录一些不错的Python相关的资料，供需要时参考：

1. [简明Python教程](#)<sup>1</sup>

Swaroop, C. H. 著，>沈洁元 译

特点：内容很全。

2. [关于python的中文文档](#)

这里：<http://www.elias.cn/Python/HomePage>介绍了很多的资源。其中就有[python教程的中文版](#)<sup>2</sup>

---

<sup>1</sup> [http://woodpecker.org.cn/abyteofpython\\_cn/chinese/](http://woodpecker.org.cn/abyteofpython_cn/chinese/)

<sup>2</sup> [http://wiki.woodpecker.org.cn/moin/March\\_Liu/PyTutorial](http://wiki.woodpecker.org.cn/moin/March_Liu/PyTutorial)

---

# 参考书目

- [1] [Python基础篇](#)<sup>1</sup>
- [2] [Big5 \(Traditional Chinese\) character code table](#)<sup>2</sup>
- [3] [简繁转换](#)<sup>3</sup>
- [4] [Unicode Lookup](#)<sup>4</sup>

---

<sup>1</sup> [http://www.tsnc.edu.cn/default/tsnc\\_wgrj/doc/python/basic.htm](http://www.tsnc.edu.cn/default/tsnc_wgrj/doc/python/basic.htm)

<sup>2</sup> <http://ash.jp/code/cn/big5tbl.htm>

<sup>3</sup> <http://www.j4.com.tw/big-gb/>

<sup>4</sup> <http://unicodelookup.com/>