

python常用标准库及三方库

作者：yegle

链接：

<https://www.zhihu.com/question/20501628/answer/15321184>

著作权归作者所有。商业转载请联系作者获得授权，非商业转载请注明出处。只介绍我用过觉得有强烈推荐欲望的几个

===== 标准库

* <http://docs.python.org/library/argparse.html>

写命令行脚本必备，可惜是 2.7 才添加的，得看具体环境了... * <http://docs.python.org/library/htmlparser.html>

解析 HTML DOM 树，偶尔搞搞命令行自动表单提交用得上。感觉不好用但是毕竟是官方库=====

第三方库* <http://pypi.python.org/pypi/sh>

简直是系统管理神器，谁用谁知道* <http://pypi.python.org/pypi/suds>

轻量的 SOAP 客户端，如果贵厂内部有用 SOAP 接口，那这个几乎是必须了* <http://pypi.python.org/pypi/requests>

这个 HTTP lib 的名字叫 Python HTTP for Humans 用过 urllib、urllib2、httplib的人，你懂的*

<http://pypi.python.org/pypi/pelican>

* <http://pypi.python.org/pypi/Markdoc>

前者是静态博客生成器 (写 markdown 文本 , 然后自动转换成 html 静态文件) , 后者是文档生成器(以前拿这个做 wiki 用) * <http://pypi.python.org/pypi/pep8>

检查 Python 脚本是否符合 PEP8 的 style guide*

<http://pypi.python.org/pypi/Pygments>

语法高亮的 lib , 很多 ruby 项目用的都是这个 python lib 来做语法高亮的其他想到了再补充吧作者 : atupal

链接 :

<https://www.zhihu.com/question/20501628/answer/19542741>

著作权归作者所有。商业转载请联系作者获得授权 , 非商业转载请注明出处。standard libs:

itertools <http://docs.python.org/2/library/itertools.html>

functools <http://docs.python.org/2/library/functools.html>

好 python 有必要掌握上面这两个库吧 , re 正则 subprocess

<http://docs.python.org/2/library/subprocess.html>调用 shell

命令的神器 pdb 调试 traceback 调试 pprint 漂亮的输出

logging 日志 threading 和 multiprocessing 多线程

urllib/urllib2/httplib http 库 , httplib 底层一点 , 推荐第三方的

库 requestsos/sys 系统 , 环境相关 Queue 队列

pickle/cPickle 序列化工具 hashlib md5, sha 等 hash 算法

cvsjson/simplejson python的 json 库 , 据 so 上的讨论和 benchmark , simplejson 的性能要高于 jsontimeit 计算代码运行的时间等等 cProfile python性能测量模块 glob 类似与 listfile , 可以用来查找文件 atexit 有一个注册函数 , 可用于正好在脚本退出运行前执行一些代码 dis python 反汇编 , 当对某条语句不理解原理时 , 可以用 dis.dis 函数来查看代码对应的 python 解释器指令等等。

3th libs:

paramiko <https://github.com/paramiko/paramiko> ssh

python 库 selenium <https://pypi.python.org/pypi/selenium>

浏览器自动化测试工具 selenium 的 python 接口 lxml

<http://lxml.de/> python解析 html,xml 的神器 mechanize

<https://pypi.python.org/pypi/mechanize/> Stateful

programmatic web browsing

pycurl <https://pypi.python.org/pypi/pycurl> cURL library

module for PythonFabric <http://docs.fabfile.org/en/1.8/>

Fabric is a Python (2.5 or higher) library and command-line tool for streamlining the use of SSH for application deployment or systems administration tasks.

xmltodict <https://github.com/martinblech/xmltodict> 转

dict ,真心好用 urllib3 和 requests: 当然其实 requests 就够了 Requests: HTTP for Humansflask

<http://flask.pocoo.org/python> web 微框架 ipdb 调试神器 ,
同时推荐 ipython ! 结合 ipython 使用 redis python 接
口 pymongo mongodbpython 接口 PIL

<http://www.pythonware.com/products/pil/> python 图像处理

mako <http://www.makotemplates.org/> python 模版引擎

numpy , scipy 科学计算 matplotlib 画图

scrapy 爬虫

django/tornado/web.py/web2py/uliweb/flask/twisted/bottle/

cherrypy.等等 python web 框架/服务器 sh 1.08 — sh v1.08

documentation 用来运行 shell 模块的 极佳选择

python 除了关键字 (keywords) 和内置的类型和函数
(builtins) ,更多的功能是通过 libraries(即 modules) 来提
供的。常用的 libraries (modules) 如下 : 1) python 运行
时服务* copy : copy 模块提供了对复合 (compound) 对象
(list , tuple , dict , custom class) 进行浅拷贝和深拷贝的
功能。* pickle : pickle 模块被用来序列化 python 的对象到
bytes 流 , 从而适合存储到文件 , 网络传输 , 或数据库存储。
(pickle 的过程也被称 serializing,marshalling 或者 flattening ,
pickle 同时可以用来将 bytes 流反序列化为 python 的对象) 。
* sys :sys 模块包含了跟 python 解析器和环境相关的变量和
函数。* 其他 : atexit , gc , inspect , marshal , traceback ,
types , warnings , weakref。 2) 数学* decimal : python 中

的 float 使用双精度的二进制浮点编码来表示的 ,这种编码导致了小数不能被精确的表示 , 例如 0.1 实际上内存中为 0.1000000000000000001 , 还有 $3 \times 0.1 == 0.3$ 为 False.

decimal 就是为了解决类似的问题的 , 拥有更高的精确度 , 能表示更大范围的数字 , 更精确地四舍五入。

* math : math 模块定义了标准的数学方法 , 例如 $\cos(x)$, $\sin(x)$ 等。

* random : random 模块提供了各种方法用来产生随机数。

* 其他 : fractions, numbers。

3) 数据结构 , 算法和代码简化

* array : array 代表数组 , 类似与 list , 与 list 不同的是只能存储相同类型的对象。

* bisect : bisect 是一个有序的 list , 其中内部使用二分法 (bitsection) 来实现大部分操作。

* collections : collections 模块包含了一些有用的容器的高性能实现 , 各种容器的抽象基类 , 和创建 name-tuple 对象的函数。例如包含了容器 deque , defaultdict, namedtuple 等。

* heapq : heapq 是一个使用 heap 实现的带有优先级的 queue。

* itertools : itertools 包含了函数用来创建有效的 iterators。所有的函数都返回 iterators 或者函数包含 iterators (例如 generators 和 generators)。

* operator : operator 提供了访问 python 内置的操作和解析器提供的特殊方法 , 例如 $x+y$ 为 $\text{add}(x, y)$, $x+=y$ 为 $\text{iadd}(x, y)$, $a \% b$ 为 $\text{mod}(a, b)$ 等等。

* 其他 : abc , contextlib , functools

4) string 和 text 处理

* codecs : codecs 模块被用来处理不同的字符编码

与 unicode text io的转化。* re : re 模块用来对字符串进行正则表达式的匹配和替换。* string : string 模块包含大量有用的常量和函数用来处理字符串。也包含了新字符串格式类。* struct : struct 模块被用来在 python 和二进制结构间实现转化。* unicodedata:unicodedata模块提供访问 unicode 字符数据库

5) python 数据库访问* 关系型数据库拥有共同的规范 Python Database API Specification V2.0, MySQL , Oracle 等都实现了此规范 , 然后增加自己的扩展。* sqlite3: sqlite3 模块提供了 SQLite 数据库访问的接口。SQLite 数据库是以一个文件或内存的形式存在的自包含的关系型数据库。* DBM-style 数据库模块 python提供了打了的 modules 来支持 UNIX DBM-style 数据库文件。dbm 模块用来读取标准的 UNIX-dbm 数据库文件 , gdbm 用来读取 GNU dbm 数据库文件 , dbhash 用来读取 Berkeley DB 数据库文件。所有的这些模块提供了一个对象实现了基于字符串的持久化的字典 , 他与字典 dict 非常相似 , 但是他的 keys 和 values 都必须是字符串。* shelve : shelve 模块使用特殊的“shelf”对象来支持持久化对象。这个对象的行为与 dict 相似 , 但是所有的他存储的对象都使用基于 hashtable的数据库 (dbhash , dbm , gdbm) 存储在硬盘。与 dbm 模块的区别是所存储的对象不仅是字符串 , 而且可以是任意的与 pickle 兼容的对象。

6) 文件和目录处理* bz2 : bz2 模块用来处理

以 bzip2 压缩算法压缩的文件。* filecmp : filecmp 模块提供了函数来比较文件和目录。* fnmatch: fnmatch 模块提供了使用 UNIX shell-style 的通配符来匹配文件名。这个模块只是用来匹配，使用 glob 可以获得匹配的文件列表。* glob : glob 模块返回了某个目录下与指定的 UNIX shell 通配符匹配的所有文件。* gzip : gzip 模块提供了类 GzipFile，用来执行与 GNUgzip 程序兼容的文件的读写。* shutil: shutil 模块用来执行更高级别的文件操作，例如拷贝，删除，改名。shutil 操作之针对一般的文件，不支持 pipes，block devices 等文件类型。* tarfile : tarfile 模块用来维护 tar 存档文件。tar 没有压缩的功能。* tempfile: tempfile 模块用来产生临时文件和文件名。* zipfile: zipfile 模块用来处理 zip 格式的文件。* zlib, zlib 模块提供了对 zlib 库的压缩功能的访问。7) 操作系统的服务* commands : commands 模块被用来执行简单的系统命令，命令以字符串的形式传入，且同时以字符串的形式返回命令的输出。但是此模块只在 UNIX 系统上可用。* configparser, configparser 模块用来读写 windows 的 ini 格式的配置文件。* datetime, datetime 模块提供了各种类型来表示和处理日期和时间。* errno, 定义了所有的 errorcode 对应的符号名字。* io, io 模块实现了各种 IO 形式和内置的 open() 函数。* logging, logging 模块灵活方便地对应用程序记录 events, errors, warnings, 和 debugging

信息。这些 log 信息可以被收集 ,过滤 ,写到文件或系统 log ,甚至通过网络发送到远程的机器上。*mmap , mmap 模块提供了内存映射文件对象的支持 ,使用内存映射文件与使用一般的文件或 byte 字符串相似。*msvcrt , mscrt 只可以在 windows 系统使用 ,用来访问 Visual C 运行时库的很多有用的功能。*optparse , optparse 模块更高级别来处理 UNIX style 的命令行选项 sys.argv。* os , os 模块对通用的操作系统服务提供了可移植的 (portable) 的接口。os 可以认为是 nt 和 posix 的抽象。nt 提供 windows 的服务接口 , posix 提供 UNIX (linux , mac) 的服务接口。* os.path , os.path 模块以可移植的方式来处理路径相关的操作。* signal , signal 模块用来实现信号 (signal) 处理 , 往往跟同步有关。* subprocess , subprocess 模块包含了函数和对象来统一创建新进程 , 控制新进程的输入输出流 , 处理进程的返回。* time , time 模块提供了各种时间相关的函数。常用的 time.sleep()。* winreg, winreg 模块用来操作 windows 注册表。

* 其他 :fcntl。 8) 线程和并行* multiprocessing , multiprocessing 模块提供通过 subprocess 来加载多个任务 , 通信 , 共享数据 , 执行各种同步操作。* threading , threading 模块提供了 thread 类很很多的同步方法来实现多线程编程。

* queue , queue 模块实现了各种多生产者 , 多消费者队列 , 被用来实现多线程程序的信息安全交换。* 其他 :Coroutines

and Microthreading 9) 网络编程和套接字 (sockets) *

asynchat , asynchat 模块通过封装 asyncore 来简化了应用程序的网络异步处理。* ssl , ssl 模块被用来使用 secure sockets layer(SSL) 包装 socket 对象 , 从而使得实现数据加密和终端认证。python 使用 openssl 来实现此模块。*

socketserver , socketserver 模块提供了类型简化了 TCP , UDP 和 UNIX 领域的 socket server 的实现。* 其他 :

asyncore , select。 10) internet 应用程序编程* ftplib , ftplib 模块实现了 ftp 的 client 端协议。此模块很少使用 , 因为 urllib 提供了更高级的接口。* http 包 , 包含了 http client 和 server 的实现和 cookies 管理的模块。* smtplib , smtplib 包含了 smtp client 的底层接口 , 用来使用 smtp 协议发送邮件。*

urllib , urllib 包提供了高级的接口来实现与 http server , ftp server 和本地文件交互的 client。* xmlrpc , xmlrpc 模块被用类实现 XML-RPC client。 11) web 编程* cgi , cgi 模块用来实现 cgi 脚本 , cgi 程序一般地被 webserver 执行 , 用来处理用户在 form 中的输入 , 或生成一些动态的内容。当与 cgi 脚本有管的 request 被提交 , webserver 将 cgi 作为子进程执行 , cgi 程序通过 sys.stdin 或环境变量来获得输入 , 通过 sys.stdout 来输出。* webbrowser , webbrowser 模块提供了平台独立的工具函数来使用 web browser 打开文档。* 其他 :

wsgiref/WSGI (Python Web Server Gateway Interface). 12)

internet 数据处理和编码* base64 , base64 模块提供了 base64 , base32 , base16 编码方式 , 用来实现二进制与文本间的编码和解码。base64 通常用来对编码二进制数据 , 从而嵌入到邮件或 http 协议中。* binascii , binascii 模块提供了低级的接口来实现二进制和各种 ASCII 编码的转化。* csv , csv 模块用来读写 comma-separated values(CSV) 文件。* email , email 包提供了大量的函数和对象来使用 MIME 标准来表示 , 解析和维护 email 消息。* hashlib , hashlib 模块实现了各种 secure hash 和 message digest algorithms, 例如 MD5 和 SHA1。* htmlparser(html.parser) , 此模块定义了 HTMLParser 来解析 HTML 和 XHTML 文档。使用此类 , 需要定义自己的类且继承于 HTMLParser。* json , json 模块被用类序列化或饭序列化 Javascript object notation (JSON) 对象。* xml,xml 包提供了各种处理 xml 的方法。

参考 : python essential 4.pdf

您的评论 *感谢支持，给文档评个星吧！

写点评论支持下文档

240

[发布评论](#)

~~暂无评论~~

评价文档：

分享到：

[QQ空间](#)[新浪微博](#) [微信](#)

扫二维码，快速分享到微信朋友圈

文档可以转存到百度网盘啦！

转为pdf格式

转为其他格式 >

VIP专享文档格式自由转换

下载券

立即下载

加入VIP

免券下载