

wget命令的使用

http://blog.chinaunix.net/uid-14735472-id-111049.html

用wget做站点镜像

wget -r -p -np -k http://xxx.edu.cn

-r 表示递归下载,会下载所有的链接,不过要注意的是,不要单独使用这个参数,因为如果你要下载的网站也有别的网站的链接,wget也会把别的网站的东西下载下来,所以要加上-np这个参数,表示不下载别的站点的链接.

-np 表示不下载别的站点的链接.

-k 表示将下载的网页里的链接修改为本地链接.

-p 获得所有显示网页所需的元素,比如图片什么的.

-E 或 --html-extension 将保存的URL的文件后缀名设定为 ".html"

+++++

wget -c -t 0 -O rhel6_x86_64.iso

http://zs.kan115.com:8080/rhel6_x86_64.iso

-c 断点续传

-t 0 反复尝试的次数, 0为不限次数

-O rhel6_x86_64.iso 把下载的文件命名为rhel6_x86_64.iso

http://zs.kan115.com:8080/rhel6_x86_64.iso 要下载的文件网址

+++++

wget高级用法

http://blog168.chinaunix.net/space.php?uid=26050273&do=blog&id=1742503

摘要: 本文讲述了wget的一些高级用法, 比如另存为, 后台下载, 断点下载. 批量下载.

增加下载尝试次数和测试下载链接是否生效.

记录下载日志, 下载和排除指定类型文件.

1、下载单个文件

wget url+filename

下载过程中同时可以看到四项信息

已经下载的比例

已经下载的大小

当前下载的速度

剩余的时间

公告

昵称: 静之深

园龄: 13年8个月

粉丝: 86

关注: 46

+加关注

<	2024年5月						>
日	一	二	三	四	五	六	
28	29	30	1	2	3	4	
5	6	7	8	9	10	11	
12	13	14	15	16	17	18	
19	20	21	22	23	24	25	
26	27	28	29	30	31	1	
2	3	4	5	6	7	8	

搜索

找找看

常用链接

我的随笔

我的评论

我的参与

最新评论

我的标签

我的标签

工作日志(1)

随笔分类

C/C++杂项(50)

.NET开发(50)

AI/机器学习(1)

android开发(11)

arm(1)

bat脚本研究(1)

cmake及编译调试方法(30)

ecos学习(8)

ffmpeg(27)

GO(4)

GUI开发&minigui等(2)

ios(2)

java(7)

javascript(4)

jQuery & bootstrap(6)

更多

随笔档案

2018年8月(1)

2018年7月(1)

2018年6月(1)

2017年12月(1)

2017年9月(3)

2017年7月(1)

2017年6月(7)

2017年5月(8)

2、使用一个大写O做参数表示另存为

wget -O save_name url

这种方法适用于对应链接中没有显式文件名的情况。

例如: wget -O xx.zip http://www.vim.org/scripts/download_script.php?src_id=7701

再用不带-O参数的下载一次。

ls -al

总计 132

drwxr-xr-x 2 root root 4096 07-12 10:43 .

drwxr-xr-x 4 root root 4096 07-11 16:26 ..

-rw-r--r-- 1 root root 50243 07-12 10:43 download_script.php?

src_id=7701

-rw-r--r-- 1 root root 50243 07-12 10:43 xx.zip

我们发现, 下载的大小都是一样。但是不带-O参数的, 文件名还要转换一次。不如用-O参数方便。

mv "download_script.php?src_id=7701" yy.zip

3、指定下载速率

方法是使用wget --limit-rate

wget程序默认是使用所有的带宽, 如果

是在生产服务器上下载很大的文件就不可接受了。

为了避免这种情况使用--limit-rate参数

wget --limit-rate=200k <http://www.openss7.org/repos/tarballs/strx25-0.9.2.1.tar.bz2>

4、断点下载

使用wget -c完成未完成的下载

下载到一半需要停下来干别的事情, 用^c就可以停顿住。

回来后, 继续下载可以加一个-c参数。

注意: 如果不加入-c, 那么下载的文件会多出一个.1的后缀。

5、在后台下载

方法: 加一个-b的参数

wget -b url/filename

为后台下载。下载经过写入到wget-log文件中。

用tail -f wget-log查看下载日志

6、模拟在浏览器下下载

有的网站不允许客户在非浏览器环境下下载。使用--user-agent来设置

2017年4月(7)

2017年2月(1)

2017年1月(9)

2016年12月(42)

2016年11月(21)

2016年10月(6)

2016年9月(12)

更多

阅读排行榜

1. socket编程——服务器遇到Broken Pipe崩溃(37758)
2. 什么是AOP?(34382)
3. H264码流解析及NALU(25302)
4. Windows中杀死占用某个端口的进程(17465)
5. libtool工具的使用(17271)

评论排行榜

1. GDB下查看内存命令(x命令)(2)
2. 网络编程中一些小函数, getifaddrs()等(2)
3. 各大网站收录、搜索引擎的提交入口(1)
4. Windows中杀死占用某个端口的进程(1)
5. 什么时候使用引用? 和什么时候使用指针(1)

推荐排行榜

1. 什么是AOP?(7)
2. Windows中杀死占用某个端口的进程(3)
3. 海康ipc onvif抓包分析(2)
4. 什么时候使用引用? 和什么时候使用指针(2)
5. 探讨read的返回值的三种情况(2)

最新评论

1. Re:各大网站收录、搜索引擎的提交入口
Backdata 搜索引擎网址提交入口
--静心克己
2. Re:POSIX semaphore: sem_open, sem_close, sem_post, sem_wait
很不理解, 为什么一篇文章不断转载, 制造垃圾? ? ? ?
--zqybegin
3. Re:GDB下查看内存命令(x命令)
vscode 里面 debug -exec x/<n/f/u> addr/var
--microestc
4. Re:GDB下查看内存命令(x命令)
好文章
--microestc
5. Re:什么时候使用引用? 和什么时候使用指针
请问为什么传递类对象时建议用引用?
--pear557

```
wget --user-agent="Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.9.0.3)
Gecko/2008092416 Firefox/3.0.3" URL-TO-DOWNLOAD
```

7、测试下载链接

方法:使用--spider

试图做计划下载时候, 需要先检查一下下载链接是否有效。

```
wget --spider DOWNLOAD-URL
```

如果返回OK, 则表示下载链接是正确的!

例如

```
wget --spider "http://ip138.com/ips.asp?ip=58.251.193.137&action=2"
Spider mode enabled. Check if remote file exists.
--2010-07-12 11:36:32-- http://ip138.com/ips.asp?
ip=58.251.193.137&action=2
正在解析主机 ip138.com... 221.5.47.136
Connecting to ip138.com|221.5.47.136|:80... 已连接。
已发出 HTTP 请求, 正在等待回应... 200 OK
长度: 7817 (7.6K) [text/html]
Remote file exists and could contain further links,
but recursion is disabled -- not retrieving.
```

8、增加尝试次数

方法: --tries=1000

如果网速有问题, 下载大文件的时候可能会发生错误,
默认wget尝试20次链接。

如果尝试75次, 可以

```
wget --tries=75 DOWNLOAD-URL
```

9、下载多个文件使用wget -i

将多个下载链接写入到一个download-file-list.txt文件中, 而后用

```
wget -i download-file-list.txt
```

10、下载整站

方法: 用--mirror参数

当你要下载一个完整站点并实现本地浏览的时候,

```
wget --mirror -p --convert-links -P ./LOCAL-DIR WEBSITE-URL
```

参数讲解:

--mirror: 设置这个参数用来建立本地镜像

-p: 下载所有html文件适合显示的元素

--convert-links: 下载完成后, 将文档链接都转换成本地的

-P ./LOCAL-DIR: 保存所有的文件和目录到指定文件夹下

11、下载时候禁止下载指定类型的文件

例如下载站点时候, 不打算下载gif动画图片。

wget --reject=gif WEBSITE-TO-BE-DOWNLOADED

12、记录下载日志

方法：使用小写字母o

```
wget -o xx.html.log -O xx.html "http://ip138.com/ips.asp?ip=58.251.193.137&action=2"
```

检查一下日志：

```
[root@localhost opt]# cat xx.html.log
--2010-07-12 11:57:22-- http://ip138.com/ips.asp?ip=58.251.193.137&action=2
```

正在解析主机 ip138.com... 221.5.47.136

Connecting to ip138.com[221.5.47.136]:80... 已连接。

已发出 HTTP 请求，正在等待回应... 200 OK

长度：7817 (7.6K) [text/html]

Saving to: `xx.html'

OK 100% 65.5K=0.1s

2010-07-12 11:57:22 (65.5 KB/s) - `xx.html' saved [7817/7817]

13、是第9条的增强版。可以限制下载容量

wget -Q5m -i FILE-WHICH-HAS-URLS

当下载的文件达到5兆的时候，停止下载。

注意：如果不是对一个文件下载链接清单，对单个文件，这个限制不会生效的。

14、和第11条正好相反，

这条技巧是讲述如何仅仅下载指定类型的文件

从一个网站中下载所有的pdf文件

```
wget -r -A.pdf http://url-to-webpage-with-pdfs/
```

15、使用wget完成ftp下载

匿名ftp下载类似于http下载

wget ftp-url即可。

如果是需要输入用户名和密码，则是

wget --ftp-user=USERNAME --ftp-password=PASSWORD DOWNLOAD-URL

+++++

WEB采集研究之 WGET 使用手册

<http://stream2080.blog.163.com/blog/static/299032852008769409368/>

WEB采集也快3年了，多多少少碰到不少问题，总结总结牢骚一下~有用人得到那更好

今天上经典工具wget的使用吧(奇怪blog的附件该怎么放?)

<http://www.gnu.org/software/wget/>

WGET 手册

适用于版本 wget 1.1

1. WGet的功能

- (1) 后台工作
- (2) 支持HTTP、HTTPS、FTP
- (3) 支持HTTP Proxy
- (4) 跟踪HTML、XHTML中的超链接
- (5) 创建远程站点为本地镜像
- (6) 转换下载后HTML文件的链接
- (7) 下载FTP文件时支持含通配符的文件名
- (8) 慢网速、不稳定网络支持
- (9) 支持IPv6

2. WGet命令 wget [option]... [URL]...

(1) URL格式

[http://host\[:port\]/directory/file](http://host[:port]/directory/file)

[ftp://host\[:port\]/directory/file](ftp://host[:port]/directory/file)

<ftp://user:password@host/path>

<http://user:password@host/path>

*可以对URL中的不安全字符进行“%编码”

<ftp://host/directory/file?type=a>

* “type=a” 表示以ASCII格式下载FTP文件

* “type=i” 表示以Binary格式下载FTP文件

(2) Option语法

使用GNU getopt处理命令行参数，每个选项都有简写形式(以“-”开头)和完整形式(以“--”开头)，如：

wget -r --tries=10 <http://fly.srk.fer.hr/> -o log

可以将不需要额外参数的选项连接在一起，如：

wget -drc URL 等效于 wget -d -r -c URL

可以用“--”来结束选项，如：

wget -o log -- -x

中的“-x”被当作URL来解析。

大多数选项因为是开关参数而不接受参数

(3) WGet基本启动选项

-V

--version

显示wget的版本号

-h

--help

打印描述wget所有选项的帮助信息

-b

--background

启动后立即转到后台执行；若未使用“-o”选项指定输出文件名，则输出重定向到

2 / 13

“wget-log”文件中

-e command

--execute command

执行.wgetrc配置文件中的命令，关于.wgetrc配置文件的详细内容见后面的“启动文

件”部分

(4) 日志记录及输入文件选项

-o logfile

--output-file=logfile

将所有信息记录到日志文件logfile中

-a logfile

--append-output=logfile

将日志内容添加到logfile尾部，而不是覆盖旧的logfile

-d

--debug

打开调试输出信息，该选项对于wget的开发者非常有用。若wget编译时未打开调试

支持，则此选项无效

-q

--quiet

关闭wget的输出

-v

--verbose

打开详细输出，显示所有变化信息，默认此选项是打开的

-nv

--non-verbose

关闭详细输出，但仍然会输出错误信息和基本信息

-i file

--input-file=file

从文件file中读URL，若这里“file” = “-”，则URL从标准输入读取，请使用“./-”来读取

文件名为“-”的文件

-F

--force-html

当URL输入是从一个文件中读取时，可以强制指定该文件为HTML文件，而不管此

文件实际是什么格式的。这样使您可以下载现有HTML文件中的相对链接到本地，

但需要添加“<base href=“url”>”到HTML中，或使用--base command-line选项。

-B URL

--base=URL

与“-F”选项协同工作，相当于添加“<base href=“url”>”到“-i”指定的文件中

(5) 下载选项

--bind-address=ADDRESS

当建立客户端TCP/IP连接时，将ADDRESS绑定到本地机器，ADDRESS可以是主机

名或IP地址，当您的主机绑定了多个IP时，该选项非常有用

-t number

--tries=number

设定网络不好时wget的重连次数，默认是20次，当number =0时代表无限次重连，

3 / 13

当遇到致命错误如 “connection refused” 或 “not found” （404），则停止自动重连

-O file

--output-document=file

将下载的文件全部连接在一起写入文件file

-nc

--no-clobber

若同一路径下存在相同文件名的文件则不再下载，这里clobber有点用词不当

-c

--continue

继续下载未完成的下载，例如

wget -c <ftp://sunsite.doc.ic.ac.uk/lS-IR.Z>

如果本地存在的lS-IR.Z这个文件的长度小于对应的远程文件，则wget将从本地文件的

长度的偏移量处开始下载远程同名文件

--progress=type

选择下载进度条的形式，合法的type有 “dot” 和 “bar” （默认）两种

-N

--timestamping

打开时戳，详见 “时戳” 部分

-S

--server-response

打印HTTP服务器发送的HTTP头部及FTP服务器发送的响应

--spider

当打开此选项，wget将工作在Web Spider模式下。Wget不下载页面，仅仅是检查这

些页面是否还存在。例如，可以使用wget来检查您的书签：

wget --spider --force-html -i bookmarks.html

-T seconds

--timeout=seconds

设定网络超时时间为seconds秒，等效于同时设定 “--dns-timeout” ， “--connect-timeout”

以及 “--read-timeout” 。默认是900秒读超时，超时相关选项仅接受十进制数值，支

持小数（如0.1秒）

--dns-timeout=seconds

设定DNS查询超时时间，默认无DNS超时时间

--connect-timeout=seconds

设定连接超时时间，默认无DNS超时时间

--read-timeout=seconds

设定读写超时时间（即进程挂起时间），该选项不直接影响整个下载过程，默认读

超时时间是900秒

waitretry=seconds

若不需要在重新获取页面时等待，而仅当下载失败时等待指定时间。在默认的wgetrc配置文件中此选项是关闭的

--random-wait

在0~2倍等待时间之间随机选择重新获取文件的等待时间，以躲过网站日志分析。

--no-proxy

不使用代理（即使定义了*_proxy环境变量）

-Q quota

--quota=quota

指定自动下载的限额，quota值是字节数；配额不影响单个文件的下载，主要用于

回归下载和URL输入是文件时

--no-dns-cache

关闭DNS查询缓冲

--restrict-file-names=mode

替换远端URL中在本机受限制的字符，例如，用%编码替换。默认情况下，wget是跳过这些字符（包括控制字符）。当设定mode为“unix”时，wget跳过字符“/”以及

ASCII值在0–31和128–159之间的控制字符，这在unix类操作系统中的默认规定；当

设定mode为“windows”时，wget跳过的字符包括“\”、“|”、“/”、“.”、“?”、“*”、“<”、“>”，以及与unix系统中相同的控制字符。在windows模式中，wget使用“+”代替“.”

以分隔本地文件名中的主机和端口；使用“@”代替“?”以分隔文件名中的查询参数。

例如：在unix模式下的URL：www.xemacs.org:4300/search.pl?input=blah，在windows

模式下则表示为：www.xemacs.org+4300/search.pl@input=blah。若要跳过控制字符，

则可以使用：--restrict-file-names=nocontrol

-4

--inet4-only

-6

--inet6-only

强制连接IPv4地址或IPv6地址

--prefer-family=IPv4/IPv6/none

指定优先考虑特定地址族，默认是IPv4

--retry-connrefused

把“connection refused”当成是暂态错误并再次尝试

--user=user

--password=password

为HTTP或FTP文件获取指定用户名密码

(6) 目录选项

-nd

--no-directories

当递归地获取文件是，不创建分层的目录

-x

--force-directories

强迫创建一个分层目录，即使该目录本地不存在

-nH

--no-host-directories

禁用主机名前缀的目录

--protocol-directories

使用协议名作为本地文件名的一个路径项

--cut-dirs=number

忽略number部分的目录成分，例如：

无选项 -> ftp.xemacs.org/pub/xemacs/

-nH -> pub/xemacs/

-nH --cut-dirs=2 -> .

无选项 -> <ftp.xemacs.org/pub/xemacs/>

--cut-dirs=1 -> <ftp.xemacs.org/xemacs/>

-P prefix

--directory-prefix=prefix

设定目录前缀为prefix。默认是 ""

(7) HTTP选项

-E

--html-extension

将保存的URL的文件后缀名设定为 ".html"

--http-user=user

--http-password=password

指定某个HTTP服务器上的用户名和密码，wget使用基本授权方案或消息摘要方案

对用户名和密码进行加密

--no-cache

禁用Server端的cache，这对从代理服务器上重新获取本地已经过期的文档非常有效。

wget默认是允许使用cache

--no-cookies

禁用cookie；wget默认允许使用cookie

--load-cookies file

在第一次的HTTP重新获取前，从txt文件file中加载cookie

--save-cookies file

退出前将cookie保存到file文件中，但不保存已过期的cookie和无有效期的

cookie（会

话cookie）

--keep-session-cookies

若设定此项，则--save-cookies file也保存会话cookie

--ignore-length

忽略HTTP中的Content-Length头，不仅仅下载Content-Length指定的长度

--header=header-line

在下载HTML文件中添加头部header-line，例如：

wget --header='Accept-Charset: iso-8859-2' \

--header='Accept-Language: hr' \

<http://fly.srk.fer.hr/>

wget --header="Host: foo.bar" <http://localhost/>意思是：

wget连接到localhost，但是在Host头部指定foo.bar

--proxy-user=user

--proxy-password=password

为代理服务器的授权指定用户名和密码，wget采用基本授权方法对用户名和密码

编

码

--referer=url

在HTTP请求中添加 "Referer: url"

--save-headers

6 / 13

保存HTTP发送的头部到文件中

-U agent-string

--user-agent=agent-string

将自己标识为agent-string

--post-data=string

--post-file=file

使用POST方法在HTTP请求中发送特定数据。注意，wget需要事先知道数据长度应

用举例：

#登录到服务器

```
wget --save-cookies cookies.txt \  
--post-data 'user=foo&password=bar' \  
http://server.com/auth.php
```

#开始“爬”网页

```
wget --load-cookies cookies.txt \  
-p http://server.com/interesting/article.php
```

(8) HTTPS (SSL/TLS)选项

--secure-protocol=protocol

选择使用的安全协议，protocol合法值包括：auto，SSLv2，SSLv3及TLSv1

--no-check-certificate

不按照有效证书授权检查服务器授权证书

--certificate=file

使用客户端保存在file中的证书

--certificate-type=type

设定客户端认证方式，默认方式是PEM以及DER（也就是ASN1）

--private-key=file

从文件中读取私钥

--private-key-type=type

设定私钥类型，合法值包括：PEM（默认）和DER

--ca-certificate=file

用指定文件对对等端（peer）进行CA（certificate authorities）认证

--ca-directory=directory

指定包含PEM格式的CA认证目录

--random-file=file

指定特定文件为伪随机数生成器的数据源

--egd-file=file

指定特定文件为EGD（Entropy Gathering Daemon）socket

(9) FTP选项

--ftp-user=user

--ftp-password=password

指定某个FTP服务器上的用户名和密码

--no-remove-listing

不删除在获取FTP文件时生成的临时的.listing文件

--no-glob

关闭FTP通配符，默认允许使用通配符

7 / 13

--no-passive-ftp

禁用被动FTP传输模式

--retr-symlinks

转换符号链接，并指向下载的文件

--no-http-keep-alive

关闭HTTP下载的“keep-alive”特性

(10) 递归获取选项

-r

--recursive

打开递归获取选项

-l depth

--level=depth

指定递归的最大深度，默认最大深度是5

--delete-after

在单一文档下载后立即删除

-k

--convert-links

下载完成后，转换页面中的链接以方便本地浏览

链接转换的两种方式：

- a. 指向已下载文件的链接，则转换为相对链接
- b. 指向未能下载成功的文件的链接，转换为“http://localhost/...”形式

-K

--backup-converted

转换链接前将原始文件备份（后缀为.orig）

-m

--mirror

打开适合与做镜像的选项（recursion、time-stamping等）

-p

--page-requisites

下载能够完全显示一个给定的HTML网页的所有文件

--strict-comments

精确解析HTML注释，默认是遇到第一个“-->”即终止解析

(11) 递归接受/拒绝选项

-A acclist

--accept acclist

-R rejlist

--reject rejlist

设定接受/拒绝递归下载以逗号分开的文件名前缀或模式列表

-D domain-list

--domains=domain-list

设定要追踪的域

--exclude-domains domain-list

指定不要求追踪的域

8 / 13

--follow-ftp

追踪HTML文档中的FTP链接，默认忽略所有FTP链接

--follow-tags=list

Wget包含一个HTML标记/属性对列表，按照list列表文件追踪HTML标记

--ignore-tags=list

不追踪list列表文件指定的HTML标记

-H

--span-hosts

允许跨越服务器进行递归下载

-L

--relative

仅追踪相对链接

-I list

--include-directories=list

指定追踪目录

-X list

--exclude-directories=list

指定不追踪的目录

-np

--no-parent

不向上追踪到父目录

3. 递归下载

(1) 遍历万维网的一部分或一台HTTP服务器或FTP服务器

(2) 追踪链接或目录结构

(3) 利用输入的URL或文件解析HTML

(4) HTTP递归下载采用广度优先算法，最大深度可以设定（默认为5）

(5) FTP递归下载采用深度优先算法，能够下载FTP服务器直到给定深度的所有数据

(6) 能够镜像FTP站点

(7) 默认按照远程站点目录结构创建本地目录

(8) 使用递归下载时注意使用延时操作

(9) 一些例子

? 下载单一网页, 使用--page-requisites选项

? 下载单一目录下的某些文件, 使用-np选项

? 下载单一目录下的所有文件, 使用-l 1选项

4. 链接追踪 如何避免下载不想要的链接

(1) 访问主机的树型扩展 (默认情况下, 每次下载仅访问一台主机)

? 扩展至任何关联主机XX-H选项

? 限制扩展至某些域XX-D选项

例如: wget -rH -Dserver.com <http://www.server.com/>

可以从所有*.server.com域下载文件

? 禁止扩展至某些域XX--exclude-domains选项

例如: wget -rH -Dfoo.edu --exclude-domains sunsite.foo.edu \

9 / 13

<http://www.foo.edu/>

从所有foo.edu域下载文件, 而禁止从sunsite.foo.edu下载文件

(2) 文件类型限制

-A acclist

--accept acclist

accept = acclist

--accept选项的参数是一个文件名后缀 (如gif或.gif) 或字符串模式列表 (如
czy*

or czyBear*196[0-9]*)

-R rejlist

--reject rejlist

reject = rejlist

--reject选项与--accept选项工作方式相同, 但逻辑相反

注意: 这两个选项不影响HTML文件下载

(3) 基于目录的限制

-I list

--include list

include_directories = list

下载目录列表中的目录

-X list

--exclude list

exclude_directories = list

禁止下载目录列表中的目录

-np

--no-parent

no_parent = on

禁止访问远端当前目录的上层

(4) 相对链接

打开-L选项可以仅下载超链接中的相对链接, 而不下绝对链接

(5) 追踪FTP链接

由于FTP与HTTP协议的不同, 所以要追踪FTP链接请使用--follow-ftp选项

5. 使用时戳对下载的文件进行增量更新

使用-S选项可以在本地保留页面对应的时戳, 进行增量下载是使用-N选项, 例

如:

初次下载网页使用: wget -S <http://www.gnu.ai.mit.edu/>

以后更新网页使用: wget -N <http://www.gnu.ai.mit.edu/>

对于HTTP协议, 时戳需要 "Last-Modified" 头部支持, 而对于FTP协议, 取决于wget

能够解析获得目录列表中包含的日期的格式

6. 启动文件 (wgetrc)

(1) Wgetrc的位置

wget启动时, 将查找全局的启动配置文件, 默认情况下, 这个文件位于:

“/usr/local/etc/wgetrc”; 若该位置不存在此文件, 则查找环境变量WGETRC指定的

位置; 若WGETRC环境变量未设定, wget会加载\$HOME/.wgetrc; 否则报错!

10 / 13

(2) Wgetrc语法

wgetrc命令行的语法非常简单: “variable = value”。wgetrc命令行语法不区分

大小写, 不区分有无下划线, 忽略空行和以“#”开头的行

(3) Wgetrc命令

wgetrc命令全集如下表:

一些说明: 合法值都列在“=”后面;

Bool值得设定可以用on/off或1/0;

某些命令使用了伪随机数;

“address”可以是主机名或以点分十进制表示的IP地址;

“n”可以是正整数或无穷大 (inf)

“string”可以是任何非空字符串

命令 说明

accept/reject = string 相当于-A/-R

add_hostdir = on/off 允许/禁止使用主机名前缀的文件名; 可以用-nH禁用

continue = on/off 允许/禁止继续完成下载部分完成的文件

background = on/off 允许/禁止后台运行; 相当于-b

backup_converted =

on/off

允许/禁止用后缀名.orig保存转换前的文件; 相当于-K

base = string 用于绝对地址于相对地址的替换; 相当于--base=string

bind_address = address 地址绑定; 相当于--bind-address=address

ca_certificate = file 设定权限认证文件; 相当于--ca-certificate=file

ca_directory = directory 设定权限认证文件的目录; 相当于--ca-directory=directory

cache = on/off 打开/关闭服务器cache; 相当于--no-cache

certificate = file 设定客户端认证文件; 相当于--certificate=file

certificate_type = string 设定认证类型; 相当于--certificate-type=string

check_certificate =

on/off

设定是否进行授权确认; 相当于--check-certificate

convert_links = on/off 是否转换非相对链接; 相当于-k

cookies = on/off 是否允许cookies; 相当于--cookies

connect_timeout = n 设定连接超时时间; 相当于--connect-timeout

cut_dirs = n 忽略远程第n层目录成分; 相当于--cut-dirs=n

debug = on/off 调试模式; 相当于-d

delete_after = on/off 下载后是否删除; 相当于--delete-after

dir_prefix = string 设定目录树的最顶级目录名; 相当于-P string

dirstruct = on/off 打开/关闭目录结构; 相当于-x 或 -nd

dns_cache = on/off 打开/关闭DNS cache; 相当于--no-dns-cache

dns_timeout = n 设定DNS超时时间; 相当于--dns-timeout

domains = string 相当于-D

dot_bytes = n 进度条使用点时, 设定每个点代表的Byte数 (默认1024)

dots_in_line = n 进度条使用点时, 设定每行的点数 (默认50)

11 / 13

dot_spacing = n 设定一串中点的个数 (默认10)

egd_file = file 设定EGD socket文件名为file; 相当于--egd-file=file

exclude_directories =

string
设定不用下载的目录；相当于-X
exclude_domains =
string
相当于--exclude-domains=string
follow_ftp = on/off 是否追踪HTML文件中的FTP链接；相当于--follow-ftp
follow_tags = string 仅追踪指定的HTML标记；相当于--follow-tags=string
force_html = on/off 是否强制将输入文件名的对应文件当成HTML文件（-F）
ftp_password = string 设定FTP密码
ftp_proxy = string 将string当作FTP下载的代理
ftp_user = string 设定FTP用户名
glob = on/off 打开或关闭成团下载；相当于--glob和--no-glob
header = string 为下载的HTTP文件添加头部；相当于--header=string
html_extension = on/off 在text/html或application/xhtml+xml类无后缀名的
文件名后添加
“.html” 后缀；相当于-E
http_keep_alive = on/off 是否保持HTTP链接活跃；相当于--no-http-keep-
alive
http_password = string 设定HTTP下载用的密码；相当于--http-
password=string
http_proxy = string 设定HTTP代理
http_user = string 设定HTTP 下载用户名；相当于--http-user=string
ignore_length = on/off 是否忽略Content-Length头部；相当于--ignore-
length
ignore_tags = string 递归下载时忽略某些标记；相当于--ignore-tags=string
include_directories =
string
设定下载时追踪的目录；相当于-l string
inet4_only = on/off 强制连接IPv4地址；相当于--inet4-only或-4
inet6_only = on/off 强制连接IPv6地址；相当于--inet6-only或-6
input = file 从文件中读取URL；相当于-i file
kill_longer = on/off 是否将超出Content-Length头部值的数据当作无效数据
limit_rate = rate 限定下载速率；相当于--limit-rate=rate
load_cookies = file 从文件中加载cookie；相当于--load-cookies file
logfile = file 设定日志文件；相当于-o file
mirror = on/off 打开/关闭镜像功能；相当于-m
netrc = on/off 是否读netrc文件
noclobber = on/off 相当于-nc
no_parent = on/off 是否允许下载当前目录的上层；相当于--no-parent
no_proxy = string 避免某些代理的加载
output_document = file 设定输出文件名；相当于-O file
page_requisites = on/off 是否下载用于完全显示页面的元素；相当于-p
passive_ftp =
on/off/always/never
更改被动FTP的设定；相当于--passive-ftp
password = string 同时设定FTP和HTTP下载密码
12 / 13
post_data = string 用POST方法发送所有HTTP请求，此请求的内容为一个字符
串；
相当于--post-data=string
post_file = file 用POST方法发送所有HTTP请求，此请求的内容为一个文件；相
当于--post-file=file
prefer_family =
IPv4/IPv6/none
设定优先考虑的地址族；相当于--prefer-family

private_key = file 设定私钥文件；相当于--private-key=file
private_key_type =
string
设定私钥类型；相当于--private-type=string
progress = string 设定进度条类型；相当于--progress=string
protocol_directories =
on/off
是否用协议名作为目录成分；相当于--protocol-directories
proxy_user = string 设定代理授权用户名；相当于--proxy-user=string
proxy_password = string 设定代理授权密码；相当于--proxy-
password=string
quiet = on/off 是否打开安静模式；相当于-q
quota = quota 设定下载限额
random_file = file 在系统无/dev/random的情况下，设定随机数生成目录
read_timeout = n 设定读超时时间；相当于--read-timeout=n
reclevel = n 设定递归下载的深度；相当于-l n
recursive = on/off 打开或关闭递归下载；相当于-r
referer = string 设定HTTP Referer，相当于--referer=string
relative_only = on/off 是否仅追踪相对链接；相当于-L
remove_listing = on/off 是否删除wget 下载过程中生成的 FTP 列表；相当于
--no-remove-listing
restrict_file_names =
unix/windows
相当于--restrict-file-names
retr_symlinks = on/off 当下在普通页面文件时，是否获取符号链接；相当于
--retr-symlinks
retry_connrefused =
on/off
是否将“connection refused”当作暂时错误；相当于
--retry-connrefused
robots = on/off 设定是否参考智能机器人的设定，默认是启用
save_cookies = file 将cookies保存到文件中；相当于--save-cookies file
secure_protocol = string 设定使用的安全协议；相当于--secure-
protocol=string
server_response = on/off 设定是否打印服务器响应；相当于-S
span_hosts = on/off 是否对主机名进行扩展；相当于-H
strict_comments = on/off 相当于--strict-comments
timeout = n 设定超时时间；相当于-T n
timestamping = on/off 设定是否打开时戳功能；相当于-N
tries = n 设定下载每个URL的最大尝试次数；相当于-t n
use_proxy = on/off 设定是否使用代理；相当于--no-proxy
user = string 同时设定下载HTTP和FTP文件的用户名
verbose = on/off 打开或关闭详细输出结果；相当于-v/-nv
13 / 13
wait = n 设定两次下载之间的间隔；相当于-w n
waitretry = n 设定下载尝试失败时等待的时间；相当于--waitretry=n
randomwait = on/off 是否设定请求之间的等待时间为随机数

一个奔跑的程序员

分类: [shell脚本及相关linux命令使用](#)

好文要顶

关注我

收藏该文

微信分享



静之深
粉丝 - 86 关注 - 46

0

0

[+加关注](#)

[升级成为会员](#)

« 上一篇: [UPNP](#)

» 下一篇: [C++中智能指针的设计和使用](#)

posted @ 2013-11-17 15:38 静之深 阅读(5203) 评论(0) 编辑 收藏 举报

会员力量, 点亮园子希望

[刷新页面](#) [返回顶部](#)

登录后才能查看或发表评论, 立即 [登录](#) 或者 [逛逛](#) [博客园首页](#)

【推荐】超值焕新月, 阿里云2核2G云服务器99元/年, 立即抢购

【推荐】园子周边第二季: 更大的鼠标垫, 没有logo的鼠标垫

【推荐】阿里云云市场联合博客园推出开发者商店, 欢迎关注

【推荐】会员力量, 点亮园子希望, 期待您升级成为园子会员



编辑推荐:

- 程序员天天 CURD, 怎么才能成长, 职业发展的思考?
- 小程序中使用 Lottie 动画 | 踩坑经验分享
- Windows下绑定线程到指定的CPU核心
- C++里也有菱形运算符?
- [Nano Framework ESP32篇] 使用 LCD 屏幕

阅读排行:

- 推荐3款程序员常用的画图工具
- C#的基于.net framework的Dll模块编程(五) - 编程手把手系列文章
- .Net 8.0 下的新RPC, IceRPC之试试的新玩法"打洞"
- 缓存中的这7个坑, 我把坑惨了!!!
- .NET CORE 完美支持AOT 的 ORM SqlSugar 教程