[wget 下载整个网站，或者特定目录](http://blog.csdn.net/memray/article/details/11750577)

分类： [Tools](http://blog.csdn.net/memray/article/category/1084868)2013-09-16 23:44 13785人阅读 [评论](http://blog.csdn.net/memray/article/details/11750577#comments)(0) [收藏](javascript:void(0);) [举报](http://blog.csdn.net/memray/article/details/11750577#report)

[wget](http://www.csdn.net/tag/wget)[下载](http://www.csdn.net/tag/%e4%b8%8b%e8%bd%bd)[工具](http://www.csdn.net/tag/%e5%b7%a5%e5%85%b7)

**转载自：**[**http://www.cnblogs.com/lidp/archive/2010/03/02/1696447.html**](http://www.cnblogs.com/lidp/archive/2010/03/02/1696447.html)

需要下载某个目录下面的所有文件。命令如下

wget -c -r -np -k -L -p [www.xxx.org/pub/path/](http://www.xxx.org/pub/path/)

在下载时。有用到外部域名的图片或连接。如果需要同时下载就要用-H参数。

wget -np -nH -r --span-hosts [www.xxx.org/pub/path/](http://www.xxx.org/pub/path/)

-c 断点续传  
-r 递归下载，下载指定网页某一目录下（包括子目录）的所有文件  
-nd 递归下载时不创建一层一层的目录，把所有的文件下载到当前目录  
-np 递归下载时不搜索上层目录，如wget -c -r [www.xxx.org/pub/path/](http://www.xxx.org/pub/path/)  
没有加参数-np，就会同时下载path的上一级目录pub下的其它文件  
-k 将绝对链接转为相对链接，下载整个站点后脱机浏览网页，最好加上这个参数  
-L 递归时不进入其它主机，如wget -c -r [www.xxx.org/](http://www.xxx.org/)   
如果网站内有一个这样的链接：   
[www.yyy.org](http://www.yyy.org/)，不加参数-L，就会像大火烧山一样，会递归下载[www.yyy.org](http://www.yyy.org/)网站  
-p 下载网页所需的所有文件，如图片等  
-A 指定要下载的文件样式列表，多个样式用逗号分隔  
-i 后面跟一个文件，文件内指明要下载的URL

还有其他的用法，我从网上搜索的，也一并写上来，方便以后自己使用。

wget的常见用法

wget的使用格式  
Usage: wget [OPTION]… [URL]…

\* 用wget做站点镜像:  
wget -r -p -np -k <http://dsec.pku.edu.cn/~usr_name/>  
# 或者  
wget -m <http://www.tldp.org/LDP/abs/html/>

\* 在不稳定的网络上下载一个部分下载的文件，以及在空闲时段下载  
wget -t 0 -w 31 -c <http://dsec.pku.edu.cn/BBC.avi> -o down.log &  
# 或者从filelist读入要下载的文件列表  
wget -t 0 -w 31 -c -B <ftp://dsec.pku.edu.cn/linuxsoft> -i filelist.txt -o   
down.log &

上面的代码还可以用来在网络比较空闲的时段进行下载。我的用法是:在mozilla中将不方便当时下载的URL链接拷贝到内存中然后粘贴到文件filelist.txt中，在晚上要出去系统前执行上面代码的第二条。

\* 使用代理下载  
wget -Y on -p -k <https://sourceforge.net/projects/wvware/>

代理可以在环境变量或wgetrc文件中设定

# 在环境变量中设定代理  
export PROXY=<http://211.90.168.94:8080/>  
# 在~/.wgetrc中设定代理  
http\_proxy = <http://proxy.yoyodyne.com:18023/>  
ftp\_proxy = <http://proxy.yoyodyne.com:18023/>

wget各种选项分类列表

\* 启动

-V, –version 显示wget的版本后退出  
-h, –help 打印语法帮助  
-b, –background 启动后转入后台执行  
-e, –execute=COMMAND   
执行`.wgetrc'格式的命令，wgetrc格式参见/etc/wgetrc或~/.wgetrc

\* 记录和输入文件

-o, –output-file=FILE 把记录写到FILE文件中  
-a, –append-output=FILE 把记录追加到FILE文件中  
-d, –debug 打印调试输出  
-q, –quiet 安静模式(没有输出)  
-v, –verbose 冗长模式(这是缺省设置)  
-nv, –non-verbose 关掉冗长模式，但不是安静模式  
-i, –input-file=FILE 下载在FILE文件中出现的URLs  
-F, –force-html 把输入文件当作HTML格式文件对待  
-B, –base=URL 将URL作为在-F -i参数指定的文件中出现的相对链接的前缀  
–sslcertfile=FILE 可选客户端证书  
–sslcertkey=KEYFILE 可选客户端证书的KEYFILE  
–egd-file=FILE 指定EGD socket的文件名

\* 下载

–bind-address=ADDRESS   
指定本地使用地址(主机名或IP，当本地有多个IP或名字时使用)  
-t, –tries=NUMBER 设定最大尝试链接次数(0 表示无限制).  
-O –output-document=FILE 把文档写到FILE文件中  
-nc, –no-clobber 不要覆盖存在的文件或使用.#前缀  
-c, –continue 接着下载没下载完的文件  
–progress=TYPE 设定进程条标记  
-N, –timestamping 不要重新下载文件除非比本地文件新  
-S, –server-response 打印服务器的回应  
–spider 不下载任何东西  
-T, –timeout=SECONDS 设定响应超时的秒数  
-w, –wait=SECONDS 两次尝试之间间隔SECONDS秒  
–waitretry=SECONDS 在重新链接之间等待1…SECONDS秒  
–random-wait 在下载之间等待0…2\*WAIT秒  
-Y, –proxy=on/off 打开或关闭代理  
-Q, –quota=NUMBER 设置下载的容量限制  
–limit-rate=RATE 限定下载输率

\* 目录

-nd –no-directories 不创建目录  
-x, –force-directories 强制创建目录  
-nH, –no-host-directories 不创建主机目录  
-P, –directory-prefix=PREFIX 将文件保存到目录 PREFIX/…  
–cut-dirs=NUMBER 忽略 NUMBER层远程目录

\* HTTP 选项

–http-user=USER 设定HTTP用户名为 USER.  
–http-passwd=PASS 设定http密码为 PASS.  
-C, –cache=on/off 允许/不允许服务器端的数据缓存 (一般情况下允许).  
-E, –html-extension 将所有text/html文档以.html扩展名保存  
–ignore-length 忽略 `Content-Length'头域  
–header=STRING 在headers中插入字符串 STRING  
–proxy-user=USER 设定代理的用户名为 USER  
–proxy-passwd=PASS 设定代理的密码为 PASS  
–referer=URL 在HTTP请求中包含 `Referer: URL'头  
-s, –save-headers 保存HTTP头到文件  
-U, –user-agent=AGENT 设定代理的名称为 AGENT而不是 Wget/VERSION.  
–no-http-keep-alive 关闭 HTTP活动链接 (永远链接).  
–cookies=off 不使用 cookies.  
–load-cookies=FILE 在开始会话前从文件 FILE中加载cookie  
–save-cookies=FILE 在会话结束后将 cookies保存到 FILE文件中

\* FTP 选项

-nr, –dont-remove-listing 不移走 `.listing'文件  
-g, –glob=on/off 打开或关闭文件名的 globbing机制  
–passive-ftp 使用被动传输模式 (缺省值).  
–active-ftp 使用主动传输模式  
–retr-symlinks 在递归的时候，将链接指向文件(而不是目录)

\* 递归下载

-r, –recursive 递归下载－－慎用!  
-l, –level=NUMBER 最大递归深度 (inf 或 0 代表无穷).  
–delete-after 在现在完毕后局部删除文件  
-k, –convert-links 转换非相对链接为相对链接  
-K, –backup-converted 在转换文件X之前，将之备份为 X.orig  
-m, –mirror 等价于 -r -N -l inf -nr.  
-p, –page-requisites 下载显示HTML文件的所有图片

\* 递归下载中的包含和不包含(accept/reject)

-A, –accept=LIST 分号分隔的被接受扩展名的列表  
-R, –reject=LIST 分号分隔的不被接受的扩展名的列表  
-D, –domains=LIST 分号分隔的被接受域的列表  
–exclude-domains=LIST 分号分隔的不被接受的域的列表  
–follow-ftp 跟踪HTML文档中的FTP链接  
–follow-tags=LIST 分号分隔的被跟踪的HTML标签的列表  
-G, –ignore-tags=LIST 分号分隔的被忽略的HTML标签的列表  
-H, –span-hosts 当递归时转到外部主机  
-L, –relative 仅仅跟踪相对链接  
-I, –include-directories=LIST 允许目录的列表  
-X, –exclude-directories=LIST 不被包含目录的列表  
-np, –no-parent 不要追溯到父目录