

wget

Linux系统下载文件工具

补充说明

wget命令 用来从指定的URL下载文件。wget非常稳定，它在带宽很窄的情况下和不稳定网络中有很强的适应性，如果是由于网络的原因下载失败，wget会不断的尝试，直到整个文件下载完毕。如果是服务器打断下载过程，它会再次联到服务器上从停止的地方继续下载。这对从那些限定了链接时间的服务器上下载大文件非常有用。

wget支持HTTP、HTTPS和FTP协议，可以使用HTTP代理。所谓的自动下载是指，wget可以在用户退出系统的之后在后台执行。这意味着你可以登录系统，启动一个wget下载任务，然后退出系统，wget将在后台执行直到任务完成，相对于其它大部分浏览器在下载大量数据时需要用户一直的参与，这省去了极大的麻烦。

用于从网络上下载资源，没有指定目录，下载资源回默认为当前目录。wget虽然功能强大，但是使用起来还是比较简单：

1. 支持断点下传功能 这一点，也是网络蚂蚁和FlashGet当年最大的卖点，现在Wget也可以使用此功能，那些网络不是太好的用户可以放心了；
2. 同时支持FTP和HTTP下载方式 尽管现在大部分软件可以使用HTTP方式下载，但是，有些时候，仍然需要使用FTP方式下载软件；
3. 支持代理服务器 对安全强度很高的系统而言，一般不会将自己的系统直接暴露在互联网上，所以，支持代理是下载软件必须有的功能；
4. 设置方便简单 可能，习惯图形界面的用户已经不是太习惯命令行了，但是，命令行在设置上其实有更多的优点，最少，鼠标可以少点很多次，也不要担心是否错点鼠标；
5. 程序小，完全免费 程序小可以考虑不计，因为现在的硬盘实在太大了；完全免费就不得不考虑了，即使网络上有很多所谓的免费软件，但是，这些软件的广告却不是我们喜欢的。

语法

```
wget [参数] [URL地址]
```

选项

启动参数：

```
-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
```

记录和输入文件参数：

```
-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

下载参数：

--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

目录参数：

-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 选项参数：

--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 选项参数：

```
-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)
```

递归下载参数：

```
-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
wget -S --spider url 不下载只显示过程
```

参数

URL[] 下载指定的URL地址。

实例

使用wget下载单个文件

```
wget http://www.jsdig.com/testfile.zip
```

以下的例子是从网络下载一个文件并保存在当前目录，在下载的过程中会显示进度条，包含（下载完成百分比，已经下载的字节，当前下载速度，剩余下载时间）。

下载并以不同的文件名保存

```
wget -O wordpress.zip http://www.jsdig.com/download.aspx?id=1080
```

wget默认会以最后一个符合/的后面的字符来命名，对于动态链接的下载通常文件名会不正确。

错误：下面的例子会下载一个文件并以名称download.aspx?id=1080保存：

```
wget http://www.jsdig.com/download?id=1
```

即使下载的文件是zip格式，它仍然以download.php?id=1080命名。

正确：为了解决这个问题，我们可以使用参数-o来指定一个文件名：

```
wget -o wordpress.zip http://www.jsdig.com/download.aspx?id=1080
```

wget限速下载

```
wget --limit-rate=300k http://www.jsdig.com/testfile.zip
```

当你执行wget的时候，它默认会占用全部可能的宽带下载。但是当你准备下载一个大文件，而你还需要下载其它文件时就有必要限速了。

使用wget断点续传

```
wget -c http://www.jsdig.com/testfile.zip
```

使用wget -c重新启动下载中断的文件，对于我们下载大文件时突然由于网络等原因中断非常有帮助，我们可以继续接着下载而不是重新下载一个文件。需要继续中断的下载时可以使用-c参数。

使用wget后台下载

```
wget -b http://www.jsdig.com/testfile.zip
```

```
Continuing in background, pid 1840.  
Output will be written to `wget-log'.
```

对于下载非常大的文件的时候，我们可以使用参数-b进行后台下载，你可以使用以下命令来察看下载进度：

```
tail -f wget-log
```

伪装代理名称下载

```
wget --user-agent="Mozilla/5.0 (Windows; U; Windows NT 6.1; en-US)  
AppleWebKit/534.16 (KHTML, like Gecko) Chrome/10.0.648.204 Safari/534.16"  
http://www.jsdig.com/testfile.zip
```

有些网站能通过根据判断代理名称不是浏览器而拒绝你的下载请求。不过你可以通过--user-agent参数伪装。

测试下载链接

当你打算进行定时下载，你应该在预定时间测试下载链接是否有效。我们可以增加--spider参数进行检查。

```
wget --spider URL
```

如果下载链接正确，将会显示：

```
Spider mode enabled. Check if remote file exists.  
HTTP request sent, awaiting response... 200 OK  
Length: unspecified [text/html]  
Remote file exists and could contain further links,  
but recursion is disabled -- not retrieving.
```

这保证了下载能在预定的时间进行，但当你给错了一个链接，将会显示如下错误：

```
wget --spider url  
Spider mode enabled. Check if remote file exists.  
HTTP request sent, awaiting response... 404 Not Found  
Remote file does not exist -- broken link!!!
```

你可以在以下几种情况下使用--spider参数：

- 定时下载之前进行检查
- 间隔检测网站是否可用
- 检查网站页面的死链接

增加重试次数

```
wget --tries=40 URL
```

如果网络有问题或下载一个大文件也有可能失败，wget默认重试20次连接下载文件。如果需要，你可以使用--tries增加重试次数。

下载多个文件

```
wget -i filelist.txt
```

首先，保存一份下载链接文件：

```
cat > filelist.txt  
url1  
url2  
url3  
url4
```

接着使用这个文件和参数-i下载。

镜像网站

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

下载整个网站到本地。

- --mirror开启镜像下载。
- -p下载所有为了html页面显示正常的文件。
- --convert-links下载后，转换成本地的链接。
- -P ./LOCAL保存所有文件和目录到本地指定目录。

过滤指定格式下载

```
wget --reject=gif ur
```

下载一个网站，但你不希望下载图片，可以使用这条命令。

把下载信息存入日志文件

```
wget -o download.log URL
```

不希望下载信息直接显示在终端而是在一个日志文件，可以使用。

限制总下载文件大小

```
wget -Q5m -i filelist.txt
```

当你想要下载的文件超过5M而退出下载，你可以使用。注意：这个参数对单个文件下载不起作用，只能递归下载时才有效。

下载指定格式文件

```
wget -r -A.pdf url
```

可以在以下情况使用该功能：

- 下载一个网站的所有图片。
- 下载一个网站的所有视频。
- 下载一个网站的所有PDF文件。

FTP下载

```
wget ftp-url  
wget --ftp-user=USERNAME --ftp-password=PASSWORD url
```

可以使用wget来完成

使用wget匿名

```
wget ftp-url
```

使用wget用户名和密码认证的

```
wget --ftp-user=USERNAME --ftp-password=PASSWORD url
```

From:

<https://rd.irust.top/> - 学习笔记

Permanent link:

<https://rd.irust.top/doku.php?id=command:wget>

Last update: **2021/10/15 14:58**

