

wget使用

1、使用wget下载单个文件

以下的例子是从网络下载一个文件并保存在当前目录

```
wget http://mirrors.163.com/.help/CentOS7-Base-163.repo
```

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）。

2、使用wget -O下载并以不同的文件名保存

wget默认会以最后一个符合"/"的后面的字符来命令，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.php?id=1080保存

wget <http://www.centos.bz/download?id=1>

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

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.centos.bz/download.php?id=1080>

3、使用wget -limit -rate限速下载

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。

```
wget -limit-rate=300k http://cn.wordpress.org/wordpress-3.1-zh_CN.zip
```

4、使用wget -c断点续传

使用wget -c重新启动下载中断的文件：

wget -c http://cn.wordpress.org/wordpress-3.1-zh_CN.zip

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。

5、使用wget -b后台下载

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。

wget -b http://cn.wordpress.org/wordpress-3.1-zh_CN.zip

Continuing in background, pid 1840.

Output will be written to `wget-log'.

你可以使用以下命令来察看下载进度

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

6、伪装代理名称下载

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。

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" 下载链接

7、使用wget -spider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

```
wget -spider URL
```

```
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

8、使用wget -tries增加重试次数

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。

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

9、使用wget -i下载多个文件

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

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

10、使用wget -mirror镜像网站

下面的例子是下载整个网站到本地。

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

-mirror:开户镜像下载

-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

11、使用wget -reject过滤指定格式下载

你想下载一个网站，但你不希望下载图片，你可以使用以下命令。

```
wget -reject=gif url
```

12、使用wget -o把下载信息存入日志文件

你不希望下载信息直接显示在终端而是在一个日志文件，可以使用以下命令：

wget -o download.log URL

13、使用wget -Q限制总下载文件大小

当你想要下载的文件超过5M而退出下载，你可以使用以下命令：

wget -Q5m -i filelist.txt

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。

14、使用wget -r -A下载指定格式文件

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

wget -r -A.pdf url

15、使用wget FTP下载

你可以使用wget来完成ftp链接的下载。

使用wget匿名ftp下载

wget ftp-url

使用wget用户名和密码认证的ftp下载

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

wget是在Linux下开发的开放源代码的软件，作者是Hrvoje Niksic，后来被移植到包括Windows在内的各个平台上。它有以下功能和特点：

(1) 支持断点下载功能；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；

wget虽然功能强大，但是使用起来还是比较简单的，基本的语法是：wget [参数列表] URL。下面就结合具体的例子来说明一下wget的用法。

1、下载整个http或者ftp站点。

wget <http://place.your.url/here>

这个命令可以将<http://place.your.url/here> 首页下载下来。使用-x会强制建立服务器上一模一样的目录，如果使用-nd参数，那么服务器上下载的所有内容都会加到本地当前目录。

wget -r <http://place.your.url/here>

这个命令会按照递归的方法，下载服务器上所有的目录和文件，实质就是下载整个网站。这个命令一定要小心使用，因为在下载的时候，被下载网站指向的所有地址同样会被下载，因此，如果这个网站引用了其他网站，那么被引用的网站也会被下载下来！基于这个原因，这个参数不常用。可以用-l number参数来指定下载的层次。例如只下载两层，那么使用-l 2。

要是您想制作镜像站点，那么可以使用 -m参数，例如：wget -m <http://place.your.url/here>

这时wget会自动判断合适的参数来制作镜像站点。此时，wget会登录到服务器上，读入robots.txt并按robots.txt的规定来执行。

2、断点续传。

当文件特别大或者网络特别慢的时候，往往一个文件还没有下载完，连接就已经被切断，此时就需要断点续传。wget的断点续传是自动的，只需要使用-c参数，例如：

```
wget -c http://the.url.of/incomplete/file
```

使用断点续传要求服务器支持断点续传。-t参数表示重试次数，例如需要重试100次，那么就写-t 100，如果设成-t 0，那么表示无穷次重试，直到连接成功。-T参数表示超时等待时间，例如-T 120，表示等待120秒连接不上就算超时。

3、批量下载。

如果有多个文件需要下载，那么可以生成一个文件，把每个文件的URL写一行，例如生成文件download.txt，然后用命令：wget -i download.txt

这样就会把download.txt里面列出的每个URL都下载下来。（如果列的是文件就下载文件，如果列的是网站，那么下载首页）

4、选择性的下载。

可以指定让wget只下载一类文件，或者不下载什么文件。例如：

```
wget -m -reject=gif http://target.web.site/subdirectory
```

表示下载<http://target.web.site/subdirectory>，但是忽略gif文件。-accept=LIST 可以接受的文件类型，-reject=LIST拒绝接受的文件类型。

5、密码和认证。

wget只能处理利用用户名/密码方式限制访问的网站，可以利用两个参数：

-http-user=USER设置HTTP用户

-http-passwd=PASS设置HTTP密码

对于需要证书做认证的网站，就只能利用其他下载工具了，例如curl。

6、利用代理服务器进行下载。

如果用户的网络需要经过代理服务器，那么可以让wget通过代理服务器进行文件的下载。此时需要在当前用户的目录下创建一个.wgetrc文件。文件中可以设置代理服务器：

```
http-proxy = 111.111.111.111:8080
```

```
ftp-proxy = 111.111.111.111:8080
```

分别表示http的代理服务器和ftp的代理服务器。如果代理服务器需要密码则使用：

```
-proxy-user=USER设置代理用户
```

```
-proxy-passwd=PASS设置代理密码
```

这两个参数。

使用参数-proxy=on/off 使用或者关闭代理。

wget还有很多有用的功能，需要用户去挖掘。

附录：

命令格式：

```
wget [参数列表] [目标软件、网页的网址]
```

-V,-version 显示软件版本号然后退出；

-h,-help显示软件帮助信息；

-e,-execute=COMMAND 执行一个“.wgetrc”命令

-o,-output-file=FILE 将软件输出信息保存到文件；

-a,-append-output=FILE将软件输出信息追加到文件；

-d,-debug显示输出信息；

-q,-quiet 不显示输出信息；

-i,-input-file=FILE 从文件中取得URL；

-t,-tries=NUMBER 是否下载次数 (0表示无穷次)
-O -output-document=FILE 下载文件保存为别的文件名
-nc, -no-clobber 不要覆盖已经存在的文件
-N,-timestamping只下载比本地新的文件
-T,-timeout=SECONDS 设置超时时间
-Y,-proxy=on/off 关闭代理

-nd,-no-directories 不建立目录
-x,-force-directories 强制建立目录

-http-user=USER设置HTTP用户
-http-passwd=PASS设置HTTP密码
-proxy-user=USER设置代理用户
-proxy-passwd=PASS设置代理密码

-r,-recursive 下载整个网站、目录 (小心使用)
-l,-level=NUMBER 下载层次

-A,-accept=LIST 可以接受的文件类型
-R,-reject=LIST拒绝接受的文件类型
-D,-domains=LIST可以接受的域名
-exclude-domains=LIST拒绝的域名
-L,-relative 下载关联链接
-follow-ftp 只下载FTP链接
-H,-span-hosts 可以下载外面的主机
-I,-include-directories=LIST允许的目录
-X,-exclude-directories=LIST 拒绝的目录

中文文档名在平常的情况下会被编码, 但是在 -cut-dirs 时又是正常的,
wget -r -np -nH -cut-dirs=3 ftp://host/test/
测试.txt

```
wget -r -np -nH -nd ftp://host/test/  
%B4%FA%B8%D5.txt  
wget "ftp://host/test/*"  
%B4%FA%B8%D5.txt
```

由於不知名的原因, 可能是为了避开特殊档名, wget 会自动将抓取档名的部分用 encode_string 处理过, 所以该 patch 就把被 encode_string 处理成 "%3A" 这种东西, 用 decode_string 还原成 ":", 并套用在目录与档案名称的部分, decode_string 是 wget 内建的函式。

```
wget -t0 -c -nH -x -np -b -m -P /home/sunny/NOD32view/ http://downloads1.kaspersky-labs.com/bases/ -o wget.log
```