

# httrack ,把你从写爬虫中解放出来

作者: [ss](#)

原文链接: <https://ld246.com/article/1461301116376>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

最近在做一些无聊的文书工作，基本就是复制粘贴，太累人了，本打算把整个网站写爬虫趴下了，记来这个东东，懒得趴了。

丢上vps去了，带上套，不久爬完了。

特此向大家推荐这个好东西，从写爬虫之中解放出来。[httrack](#) 跨平台，开源。官方介绍如下：

HTTrack is a free (GPL, libre/free software) and easy-to-use offline browser utility.

It allows you to download a World Wide Web site from the Internet to a local directory, building recursively all directories, getting HTML, images, and other files from the server to your computer. HTTrack arranges the original site's relative link-structure. Simply open a page of the "mirrored" website in your browser, and you can browse the site from link to link, as if you were viewing it online. HTTrack can also update an existing mirrored site, and resume interrupted downloads. HTTrack is fully configurable, and has an integrated help system.

WinHTTrack is the Windows 2000/XP/Vista/Seven release of HTTrack, and WebHTTrack the Linux/Unix/BSD release. See the download page.

一个中型网站几分钟就可以趴下了，效率是wget的上百倍。注意别玩的太狠了。