**[Wget创建一个网站的镜像](https://fosswire.com/post/2008/04/create-a-mirror-of-a-website-with-wget/)**

* 2008年4月21日
* 阿凡达为彼得

[彼得  
Upfold](https://fosswire.com/users/peter/)

* [40条评论](https://fosswire.com/post/2008/04/create-a-mirror-of-a-website-with-wget/#comments)

GNU的wget命令行程序下载是非常受欢迎的，而不是没有理由。虽然您可以使用它来简单地从服务器检索单个文件，它比这更强大，并提供了更多的功能。

wget中更高级的功能之一就是镜像功能。这样可以创建一个网站的完整本地副本，包括任何样式表，支持图片和其他支持文件。所有（内部）链接也将被跟踪和下载（及其资源），直到您在本地计算机上具有该站点的完整副本。

最基本的形式是使用镜像功能，如：

$ wget -m *http://www.example.com/*

但是，您可能会遇到几个问题。

首先，这对于本地浏览不是很有用，因为页面中的链接本身仍然指向真实的URL，而不是您的本地下载。这意味着，如果说，您下载了**http://www.example.com/**，该页面上的**http://www.example.com/page2.html**链接仍将指向example.com的服务器如果您因为某些原因在离线时尝试浏览本地的本地副本，那么将会是一个正确的痛苦。

要解决这个问题，可以使用**-k**选项和镜像选项：

$ wget -mk *http://www.example.com/*

现在，我之前提到的那个链接将指向相对的**page2.html**。所有图像，样式表和资源都会发生这种情况，因此您现在应该可以获得真实的离线浏览体验。

还有另外一个主要的问题我还没有在这里覆盖 - 带宽。无论您在连接上使用的带宽如何拉低整个站点，您将会对远程服务器造成压力。你应该考虑到善良和减少他们的负担（和你），特别是如果网站很小，带宽是非常重要的。玩得很好

您可以做到这一点的一个方法是通过在服务器的请求之间延迟来故意减慢下载速度。

$ wget -mk -w 20 *http://www.example.com/*

这在请求之间延迟20秒。替换该号码，可选地，您可以为分钟添加**m**的后缀，**h**表示几个小时，**d**表示...，是的，如果您想进一步减慢镜像速度。

现在，如果你想备份一些东西，或者下载你最喜爱的网站，以便在离线的时候进行查看，可以使用wget的镜像功能。要进一步深入了解，请查看wget的man页面（**man wget**），其中还有其他选项，如随机延迟，设置自定义用户代理，向站点发送cookies等等。