



Principales options d'utilisation de httrack sous Linux

HTTrack est un outil en ligne de commande permettant de récupérer (aspirer) un site web pour une consultation hors ligne. La syntaxe générale est :

```
httrack <URLs> [-options] [+filtres] [-filtres]
```

Modes d'action

Option courte	Option longue	Description
-w	--mirror	Mode miroir par défaut (recopie le site)
-W	--mirror-wizard	Mode interactif avec questions
-g	--get-files	Télécharge les fichiers sans recréer la structure
-i	--continue	Reprend un téléchargement interrompu
-Y	--mirrorlinks	Suit aussi les liens externes de premier niveau

(HTTrack Users Guide, Debian manpage) ^[1] ^[2]

Chemin de sortie

Option	Description
-O path	Répertoire de destination du miroir et des fichiers de cache/log

Limites de téléchargement

Option courte	Option longue	Description
-rN	--depth=N	Profondeur de liens maximale (défaut : illimitée, r9999)
%eN	--ext-depth=N	Profondeur pour les liens externes (défaut : 0)
-mN	--max-files=N	Nombre maximal de fichiers non-HTML
-MN	--max-size=N	Taille totale maximale téléchargeable
-EN	--max-time=N	Durée maximale en secondes
-AN	--max-rate=N	Débit maximal en octets/seconde

(Linux Command Library, HTTrack Users Guide) ^[3] ^[1]

Contrôle du flux et connexions

Option courte	Option longue	Description
-cN	--sockets=N	Nombre de connexions simultanées (défaut : 8)
-TN	--timeout=N	Délai d'attente en secondes avant abandon
-RN	--retries=N	Nombre de tentatives en cas d'erreur
-JN	--min-rate=N	Débit minimum toléré (anti-engorgement)
-HN	--host-control=N	Abandon d'un hôte (0=jamais, 1=timeout, 2=lent, 3=les deux)
%cN	--connection-per-second=N	Connexions max par seconde (défaut : 10)

(Debian manpage) ^[2]

Portée du crawl (scope)

Option courte	Option longue	Description
-a	--stay-on-same-address	Restreint à la même adresse (hôte)
-d	--stay-on-same-domain	Restreint au même domaine principal
-l	--stay-on-same-tld	Restreint au même TLD (ex : .com)
-e	--go-everywhere	Suit les liens vers n'importe quel hôte
-n	--near	Télécharge aussi les fichiers non-HTML « proches »
-t	--test	Teste les liens hors scope sans les télécharger

(HTTrack command-line guide) ^[4]

Filtres d'URL

Les filtres permettent d'inclure (+) ou d'exclure (-) des URLs selon des motifs :

```
+*.pdf +*.jpg -*      # uniquement les PDF et JPG
+example.com/* -*     # uniquement ce domaine, exclure tout le reste
```

(Linux Command Library) ^[3]

Proxy et authentification

Option courte	Option longue	Description
-P host:port	--proxy	Utiliser un proxy (avec auth : user:pass@proxy:port)
%f	--httpproxy-ftp	Utiliser le proxy pour FTP
-F string	—	Définir le User-Agent

Exemples pratiques

```
# Miroir simple d'un site
httrack https://example.com/ -O /home/user/mirror

# Limiter la profondeur à 3 niveaux
httrack https://example.com/ -r3 -O /home/user/mirror

# Reprendre un téléchargement interrompu
httrack --continue

# Limiter à 100 KB/s et 4 connexions
httrack https://example.com/ -A102400 -c4 -O /home/user/mirror

# Filtrer uniquement les PDF
httrack https://example.com/ +*.pdf -* -O /home/user/mirror

# Ignorer robots.txt
httrack https://example.com/ -s0 -O /home/user/mirror
```

Installation

```
sudo apt install httrack      # Debian/Ubuntu
sudo dnf install httrack      # Fedora
```

La page de manuel `man httrack` donne la liste complète et à jour des options acceptées par la version installée.^[5]



1. <https://www.httrack.com/HelpHtml/fcguide.html>
2. <https://manpages.debian.org/bookworm/httrack/httrack.1.en.html>
3. <https://linuxcommandlibrary.com/man/httrack>
4. <https://www.httrack.com/html/cmdguide.html>
5. <https://www.httrack.com/html/>
6. <https://www.httrack.com/HelpHtml/faq.html>
7. <https://www.httrack.com/src/htsalias.c.html>
8. <https://www.httrack.com/html/httrack.man.html>
9. <https://www.httrack.com/html/fcguide.html>
10. <https://manpages.debian.org/unstable/httrack/httrack.1.en.html>
11. <https://manpages.debian.org/stretch/httrack/httrack.1.en.html>