

R V COLLEGE OF ENGINEERING

Name: Dhanush M USN: 1RV18IS007 Dept/Lab: ISE/CSDF Expt No.: 02 b
Date: 26/11/2021 Title: WEB APPLICATION ANALYSIS TOOLS

b. HTTRACK - Offline Browser

INTRODUCTION

Penetration testing is a simulated cyber attack where professional ethical hackers break into corporate networks to find weaknesses even before attackers do. Httrack is one of the tools that could be used for this purpose.

HTTrack is an easy-to-use offline browser utility. It allows you to download a World Wide website from the Internet to a local directory, building recursively all directories, getting html, images, and other files from the server to your computer.

By default, 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.

HTTrack uses a Web crawler to download a website. Some parts of the website may not be downloaded by default due to the robots exclusion protocol unless disabled during the program. HTTrack can follow links that are generated with basic JavaScript and inside Applets or Flash, but not complex links (generated using functions or expressions) or server-side image maps.

Objectives - To clone a World Wide Web site from the internet to a local computer.

EXECUTION STEPS

1. Installing HTTrack from a package

Command - `sudo apt-get install httrack webhttrack`

2. Running HTTrack GUI

Command - `webhttrack`

3. Basic structure

Syntax - `httrack <URLs> [-option] [+<URL_filter>] [-<URL_filter>] [+<mime:MIME_filter>] [-<mime:MIME_filter>]`

Examples cases

1. To mirror a site

Syntax - `httrack <site_URL>`

`httrack www.someweb.com/bob/`

```
└─$ httrack www.someweb.com/bob/
Mirror launched on Thu, 02 Dec 2021 23:50:04 by HTTrack Website Copier/3.49-2+libhtsjava.so.2 [XR6C0'2014]
mirroring www.someweb.com/bob/ with the wizard help..
Done.www.someweb.com/bob/ (0 bytes) - -5
Thanks for using HTTrack!
```

2. To mirror two sites together (with shared links) and accept any .jpg files on .com sites
Syntax - `httrack <site1_URL> <site2_URL> +<mime:MIME_filter> -<mime:MIME_filter>`
`httrack www.someweb.com/bob/ www.anothertest.com/mike/ +*.com/*.jpg`
`-mime:application/*`

```
L$ httrack www.someweb.com/bob/ www.anothertest.com/mike/ +*.com/*.jpg -mime:application/*
Mirror launched on Thu, 02 Dec 2021 23:53:22 by HTTrack Website Copier/3.49-2+libhtsjava.so.2 [XR&CO'2014]
]
mirroring www.someweb.com/bob/ www.anothertest.com/mike/ +*.com/*.jpg -mime:application/* with the wizard help..
Done.www.someweb.com/bob/ (0 bytes) - -5
Thanks for using HTTrack!
```

3. To get all files starting from bobby.html, with 6 link-depth, and possibility of going everywhere on the web
Command - `httrack www.someweb.com/bob/bobby.html +* -r6`

```
L$ httrack www.someweb.com/bob/bobby.html +* -r6
There is an index.html and a hts-cache folder in the directory
A site may have been mirrored here, that could mean that you want to update it
Be sure parameters are ok

Press <Y><Enter> to confirm, <N><Enter> to abort
y
Mirror launched on Thu, 02 Dec 2021 23:58:46 by HTTrack Website Copier/3.49-2+libhtsjava.so.2 [XR&CO'2014]
]
mirroring www.someweb.com/bob/bobby.html +* with the wizard help..
Done.www.someweb.com/bob/bobby.html (0 bytes) - -5
Thanks for using HTTrack!
```

4. To run on spider using a proxy
Syntax - `httrack <site_URL> --spider -P <proxy_link>`
`httrack www.someweb.com/bob/bobby.html --spider -P proxy.myhost.com:8080`

```
L$ httrack www.someweb.com/bob/bobby.html --spider -P proxy.myhost.com:8080
There is an index.html and a hts-cache folder in the directory
A site may have been mirrored here, that could mean that you want to update it
Be sure parameters are ok

Press <Y><Enter> to confirm, <N><Enter> to abort
y
Mirror launched on Fri, 03 Dec 2021 00:01:10 by HTTrack Website Copier/3.49-2+libhtsjava.so.2 [XR&CO'2014]
]
mirroring www.someweb.com/bob/bobby.html with the wizard help..
Done.www.someweb.com/bob/bobby.html (0 bytes) - -5
Thanks for using HTTrack!
```

5. To update a mirror in the current folder
Syntax - `httrack --update`

```
L$ httrack --update
Mirror launched on Fri, 03 Dec 2021 00:04:15 by HTTrack Website Copier/3.49-2+libhtsjava.so.2 [XR&CO'2014]
]
mirroring www.someweb.com/bob/bobby.html +* with the wizard help..
Done.www.someweb.com/bob/bobby.html (0 bytes) - -5
Thanks for using HTTrack!
```

6. To bring to an interactive mode
Syntax - `httrack`

```

└─$ httrack
A cache (hts-cache/) has been found in the directory
That means you can update faster the remote site(s)
OK to Update httrack httrack?

Press <Y><Enter> to confirm, <N><Enter> to abort
y
Mirror launched on Fri, 03 Dec 2021 00:06:50 by HTTrack Website Copier/3.49-2+libhtsjava.so.2 [XR&CO'2014]
]
mirroring www.someweb.com/bob/bobby.html +* with the wizard help..
Done.www.someweb.com/bob/bobby.html (0 bytes) - -5
Thanks for using HTTrack!

```

7. To continue a mirror in the current folder

Syntax - *httrack --continue*

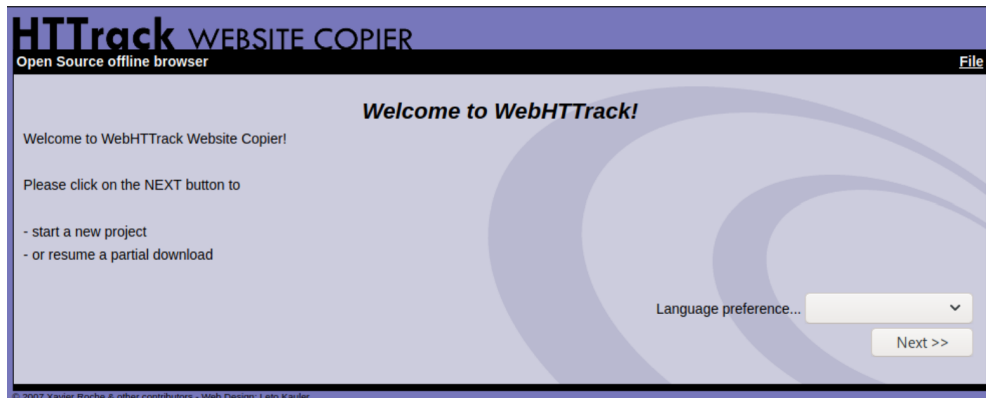
```

└─$ httrack --continue
Mirror launched on Fri, 03 Dec 2021 00:08:23 by HTTrack Website Copier/3.49-2+libhtsjava.so.2 [XR&CO'2014]
]
mirroring www.someweb.com/bob/bobby.html +* with the wizard help..
Done.www.someweb.com/bob/bobby.html (0 bytes) - -5
Thanks for using HTTrack!

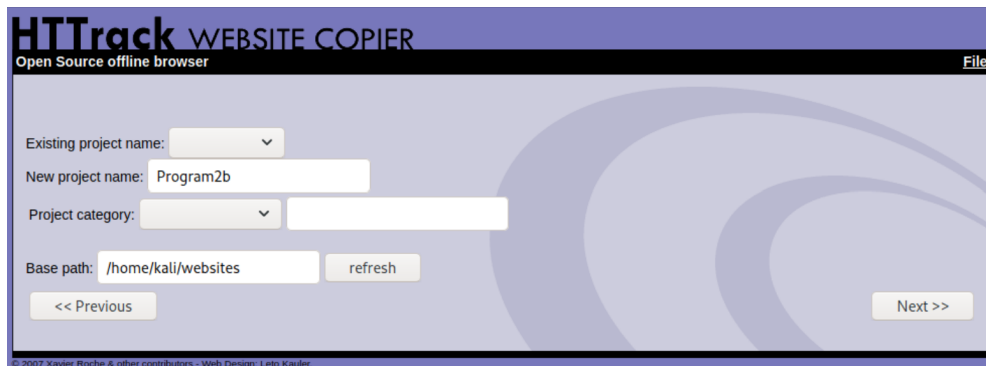
```

Using HTTrack GUI

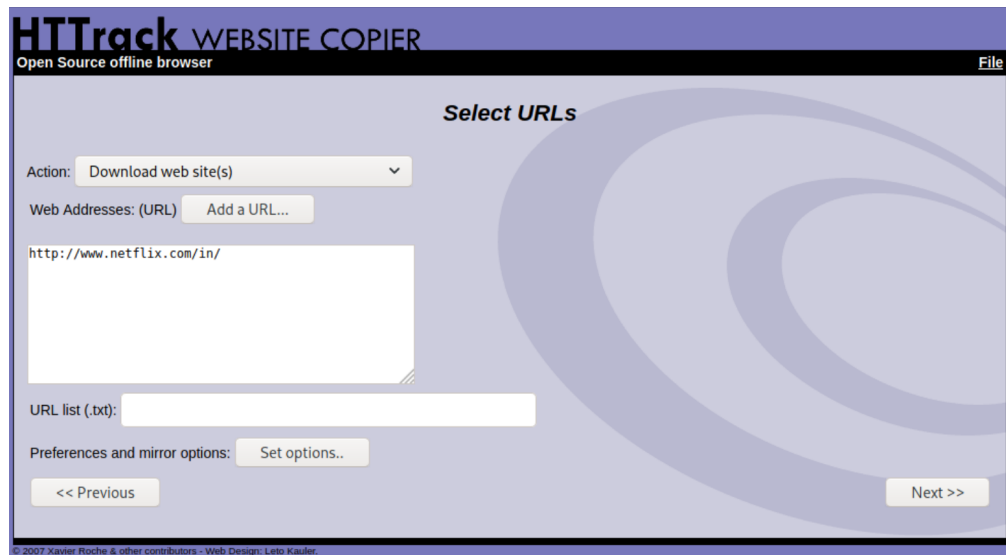
- a. Syntax - *webhttrack*



- b. Create a project



- c. Add the URL



- d. Site mirrored



CONCLUSION

1. HTTrack Website Copier is a free open source Web crawler and offline browser software program that allows you to download a copy of a World Wide Web site.
2. It can be a useful tool to make quick backups of a business's websites with its structure intact.

REFERENCES

1. HTTrack - <https://en.wikipedia.org/wiki/HTTrack>
2. How to use the HTTrack Website copier - <https://smallbusiness.chron.com/use-htrack-website-copier-52794.html>
3. HTTrack: Penetration Testing Tools - <https://en.kali.tools/?p=443>