Project Title

Web File Retrieval and Website Mirroring with `wget`

Objective

To learn and demonstrate the use of `wget` for downloading files, websites, and performing deep web retrieval, including ethical and technical considerations in cybersecurity contexts.

Skills Learned

- HTTP and web content download mechanisms
- Saving websites with full assets (CSS, JS, images)
- Recursive website mirroring
- Customizing HTTP request headers (user-agent manipulation)
- Understanding legal and ethical implications of scraping

Tools Used

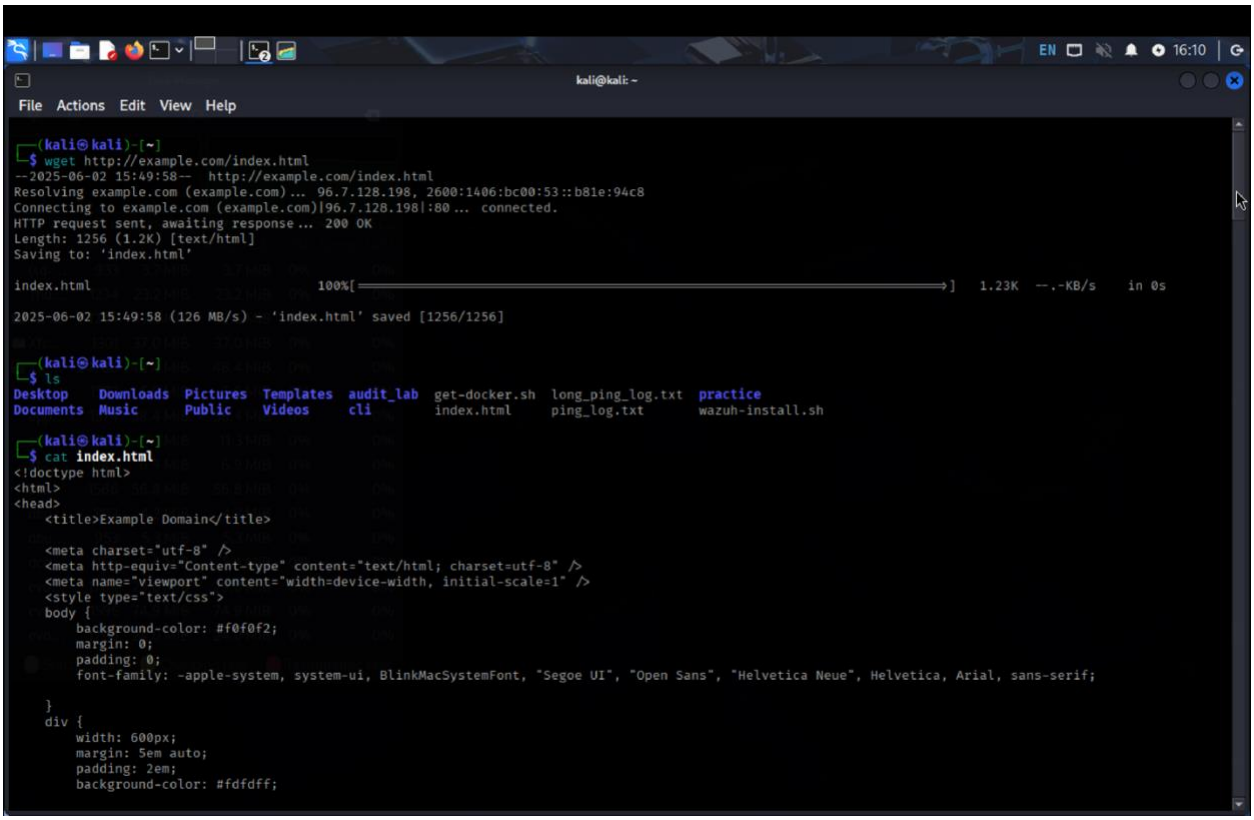
- **Operating System:** Kali Linux (or any Linux-based OS)
- **Command-Line Utility:** `wget`

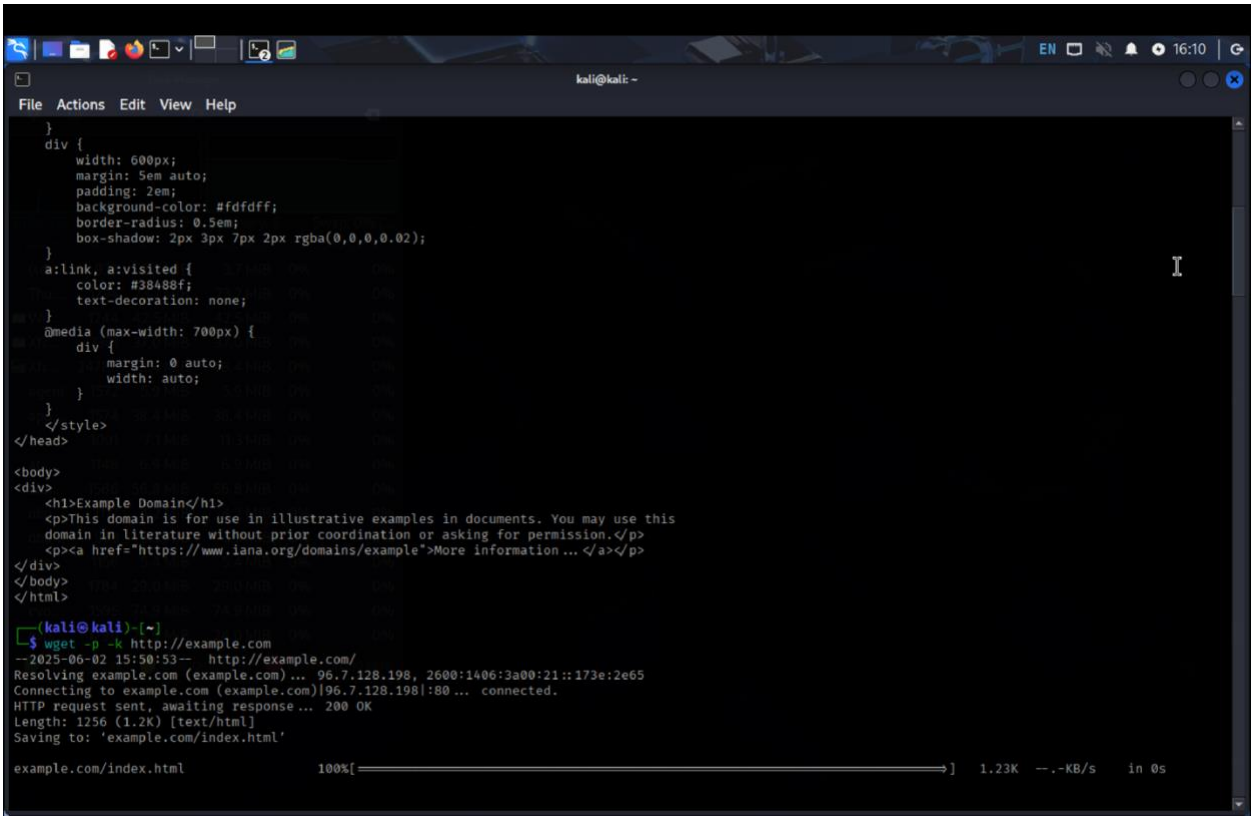
Tasks Performed

Level	Description
🟢 **Basic**	Downloaded a single file from `example.com` using `wget <URL>`.
🟡 **Intermediate**	Retrieved a full webpage with all associated resources (HTML, CSS, JS) using `wget -p -k <URL>`.
🔴 **Advanced**	Mirrored an entire website recursively using `wget -r <URL>`. Explored the risks of mass-downloading web content.
🔴 **Advanced**	Used `wget --user-agent="CustomAgent"` to spoof identity. Analyzed ethical considerations and website defenses.

Why This Matters (Real-World Impact)

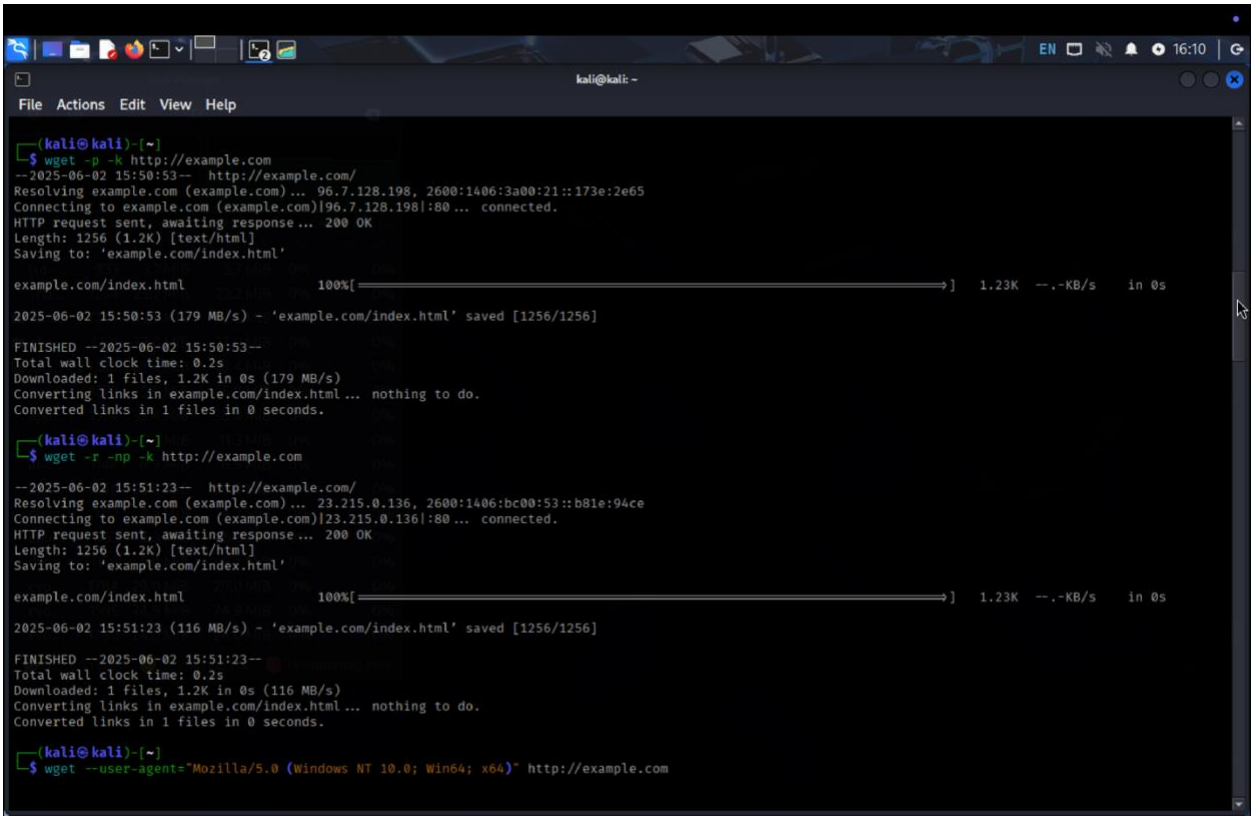
SOC analysts and cybersecurity professionals often encounter scenarios where they need to gather web-based intel, preserve malicious pages for analysis, or understand web server behavior. Tools like `wget` equip professionals to perform web reconnaissance, site cloning for phishing investigations, or forensic backups—while staying within legal boundaries.

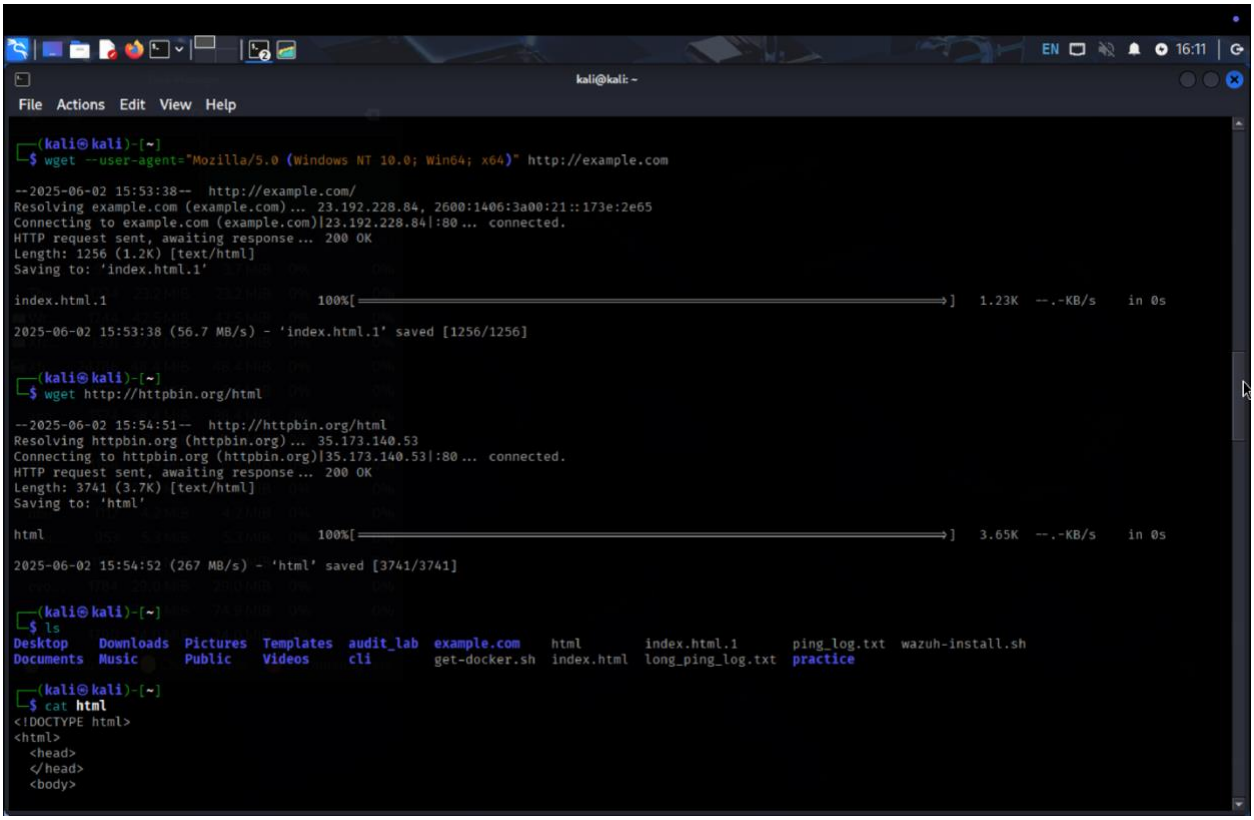

kali@kali: ~
File Actions Edit View Help
(kali㉿kali)-[~]
$ wget http://example.com/index.html
--2025-06-02 15:49:58-- http://example.com/index.html
Resolving example.com (example.com)... 96.7.128.198, 2600:1406:bc00:53::b81e:94c8
Connecting to example.com (example.com)|96.7.128.198|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1256 (1.2K) [text/html]
Saving to: 'index.html'
index.html 100%[=>] 1.23K --.-KB/s in 0s
2025-06-02 15:49:58 (126 MB/s) - 'index.html' saved [1256/1256]
(kali㉿kali)-[~]
$ ls
Desktop Downloads Pictures Templates audit_lab get-docker.sh long_ping_log.txt practice
Documents Music Public Videos cli index.html ping_log.txt wazuh-install.sh
(kali㉿kali)-[~]
$ cat index.html
<!doctype html>
<html>
<head>
<title>Example Domain</title>
<meta charset="utf-8" />
<meta http-equiv="Content-type" content="text/html; charset=utf-8" />
<meta name="viewport" content="width=device-width, initial-scale=1" />
<style type="text/css">
body {
background-color: #f0f0f2;
margin: 0;
padding: 0;
font-family: -apple-system, system-ui, BlinkMacSystemFont, "Segoe UI", "Open Sans", "Helvetica Neue", Helvetica, Arial, sans-serif;
}
div {
width: 600px;
margin: 5em auto;
padding: 2em;
background-color: #fdfdff;

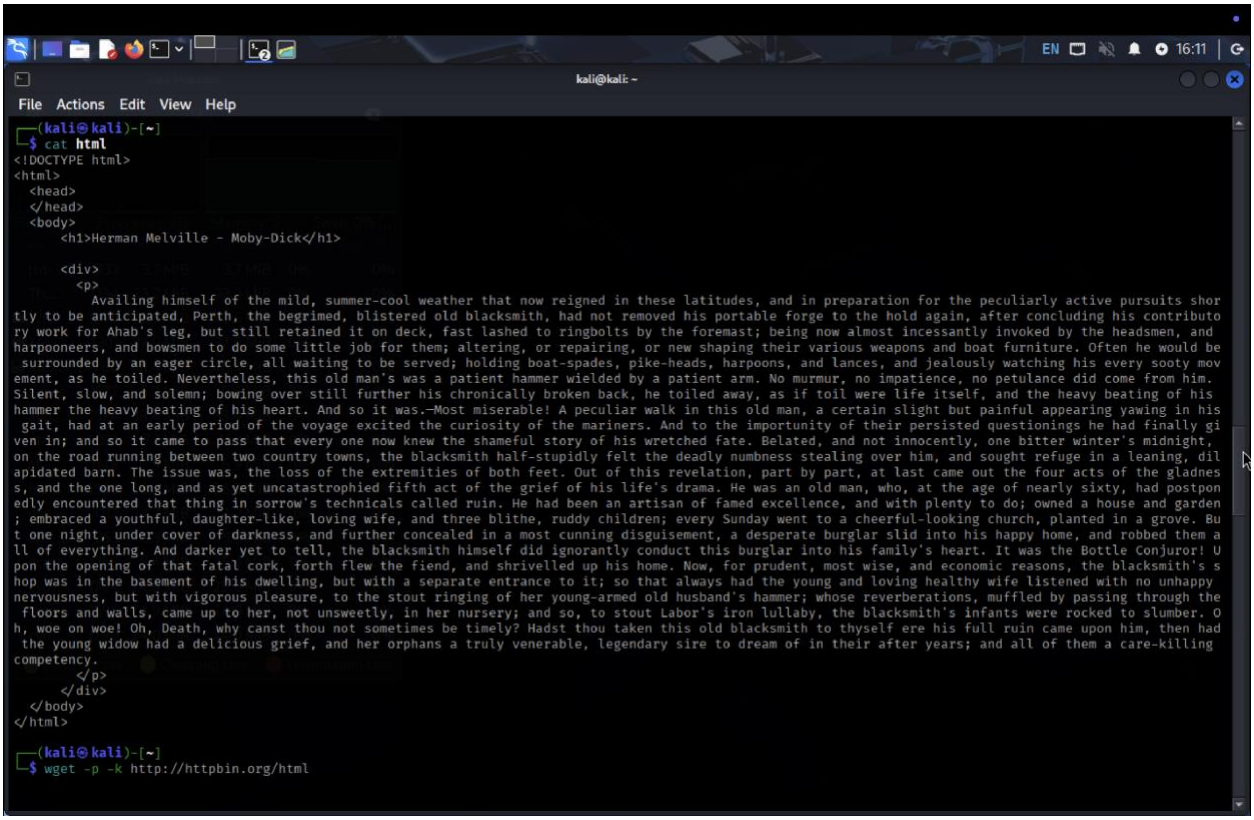

EN
16:10
kali@kali: ~
File Actions Edit View Help
}
div {
width: 600px;
margin: 5em auto;
padding: 2em;
background-color: #fdfdff;
border-radius: 0.5em;
box-shadow: 2px 3px 7px 2px rgba(0,0,0,0.02);
}
a:link, a:visited {
color: #38488f;
text-decoration: none;
}
@media (max-width: 700px) {
div {
margin: 0 auto;
width: auto;
}
}
</style>
</head>
<body>

<h1>Example Domain</h1>
This domain is for use in illustrative examples in documents. You may use this
domain in literature without prior coordination or asking for permission.
More information...

</body>
</html>
┌──(kali㉿kali)-[~]
└─$ wget -p -k http://example.com
--2025-06-02 15:50:53-- http://example.com/
Resolving example.com (example.com)... 96.7.128.198, 2600:1406:3a00:21::173e:2e65
Connecting to example.com (example.com)|96.7.128.198|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1256 (1.2K) [text/html]
Saving to: 'example.com/index.html'
example.com/index.html 100%[===>] 1.23K --.-KB/s in 0s

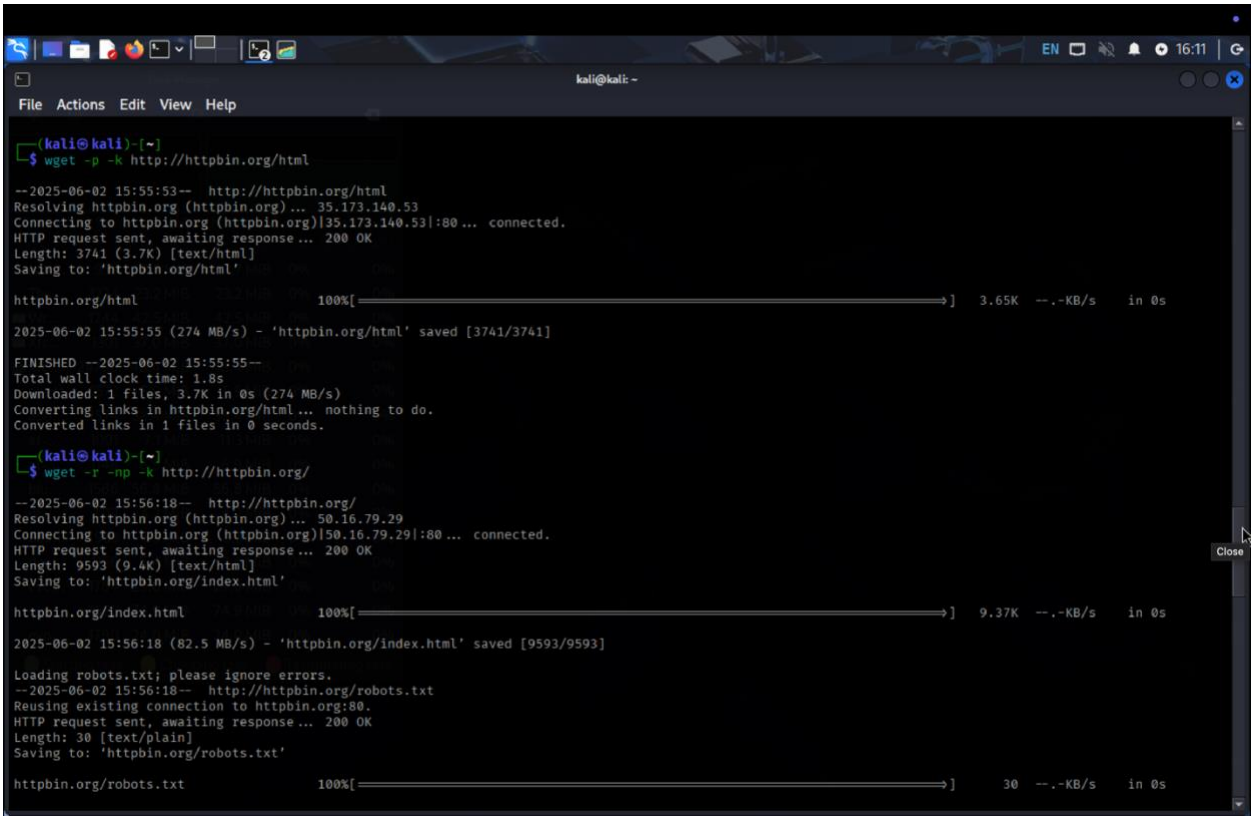

EN
16:10
kali@kali: ~
File Actions Edit View Help
┌──(kali㉿kali)-[~]
└─$ wget -p -k http://example.com
--2025-06-02 15:50:53-- http://example.com/
Resolving example.com (example.com)... 96.7.128.198, 2600:1406:3a00:21::173e:2e65
Connecting to example.com (example.com)|96.7.128.198|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1256 (1.2K) [text/html]
Saving to: 'example.com/index.html'
example.com/index.html 100%[=>] 1.23K --.-KB/s in 0s
2025-06-02 15:50:53 (179 MB/s) - 'example.com/index.html' saved [1256/1256]
FINISHED --2025-06-02 15:50:53--
Total wall clock time: 0.2s
Downloaded: 1 files, 1.2K in 0s (179 MB/s)
Converting links in example.com/index.html... nothing to do.
Converted links in 1 files in 0 seconds.
┌──(kali㉿kali)-[~]
└─$ wget -r -np -k http://example.com
--2025-06-02 15:51:23-- http://example.com/
Resolving example.com (example.com)... 23.215.0.136, 2600:1406:bc00:53::b81e:94ce
Connecting to example.com (example.com)|23.215.0.136|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1256 (1.2K) [text/html]
Saving to: 'example.com/index.html'
example.com/index.html 100%[=>] 1.23K --.-KB/s in 0s
2025-06-02 15:51:23 (116 MB/s) - 'example.com/index.html' saved [1256/1256]
FINISHED --2025-06-02 15:51:23--
Total wall clock time: 0.2s
Downloaded: 1 files, 1.2K in 0s (116 MB/s)
Converting links in example.com/index.html... nothing to do.
Converted links in 1 files in 0 seconds.
┌──(kali㉿kali)-[~]
└─$ wget --user-agent="Mozilla/5.0 (Windows NT 10.0; Win64; x64)" http://example.com

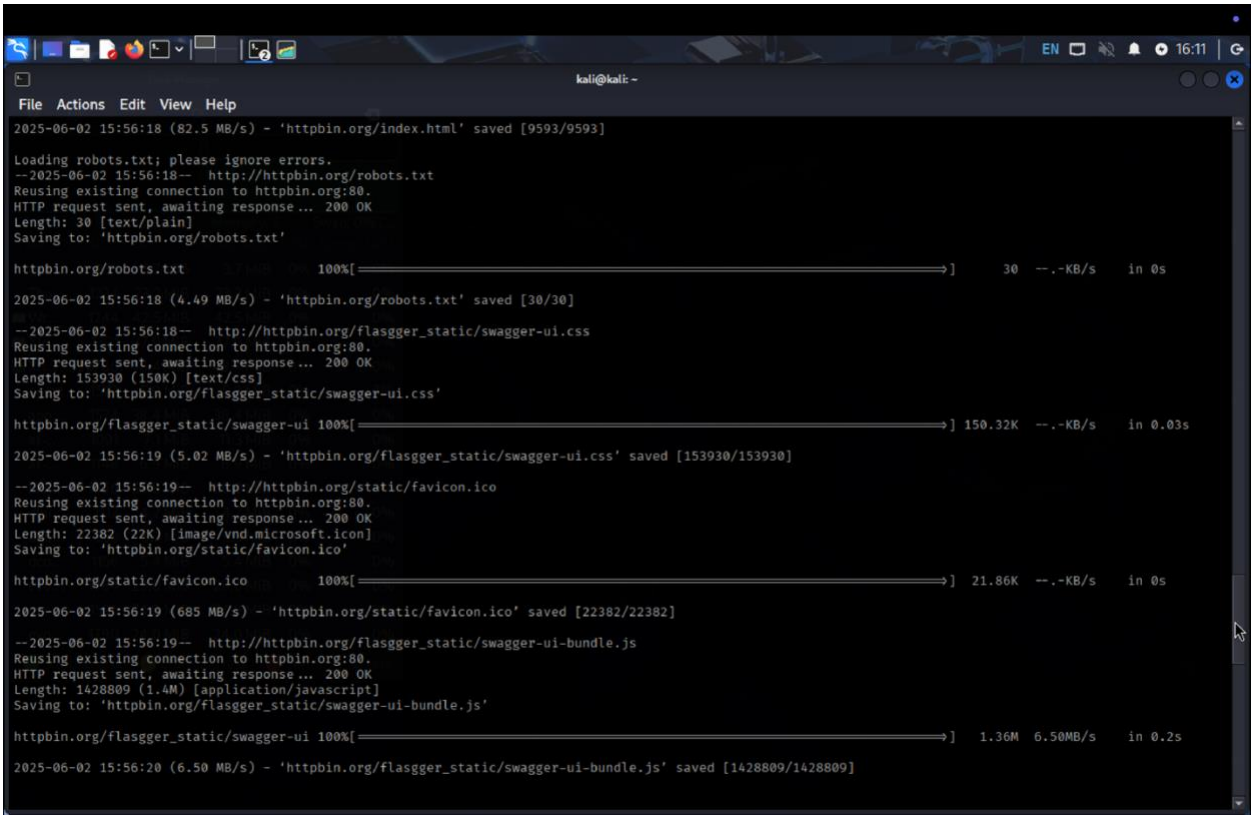

kali@kali: ~
File Actions Edit View Help
(kali㉿kali)-[~]
$ wget --user-agent="Mozilla/5.0 (Windows NT 10.0; Win64; x64)" http://example.com
--2025-06-02 15:53:38-- http://example.com/
Resolving example.com (example.com)... 23.192.228.84, 2600:1406:3a00:21::173e:2e65
Connecting to example.com (example.com)|23.192.228.84|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1256 (1.2K) [text/html]
Saving to: 'index.html.1'
index.html.1 100%[=>] 1.23K --.-KB/s in 0s
2025-06-02 15:53:38 (56.7 MB/s) - 'index.html.1' saved [1256/1256]
(kali㉿kali)-[~]
$ wget http://httpbin.org/html
--2025-06-02 15:54:51-- http://httpbin.org/html
Resolving httpbin.org (httpbin.org)... 35.173.140.53
Connecting to httpbin.org (httpbin.org)|35.173.140.53|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 3741 (3.7K) [text/html]
Saving to: 'html'
html 100%[=>] 3.65K --.-KB/s in 0s
2025-06-02 15:54:52 (267 MB/s) - 'html' saved [3741/3741]
(kali㉿kali)-[~]
$ ls
Desktop Downloads Pictures Templates audit_lab example.com html index.html.1 ping_log.txt wazuh-install.sh
Documents Music Public Videos cli get-docker.sh index.html long_ping_log.txt practice
(kali㉿kali)-[~]
$ cat html
<!DOCTYPE html>
<html>
<head>
</head>
<body>

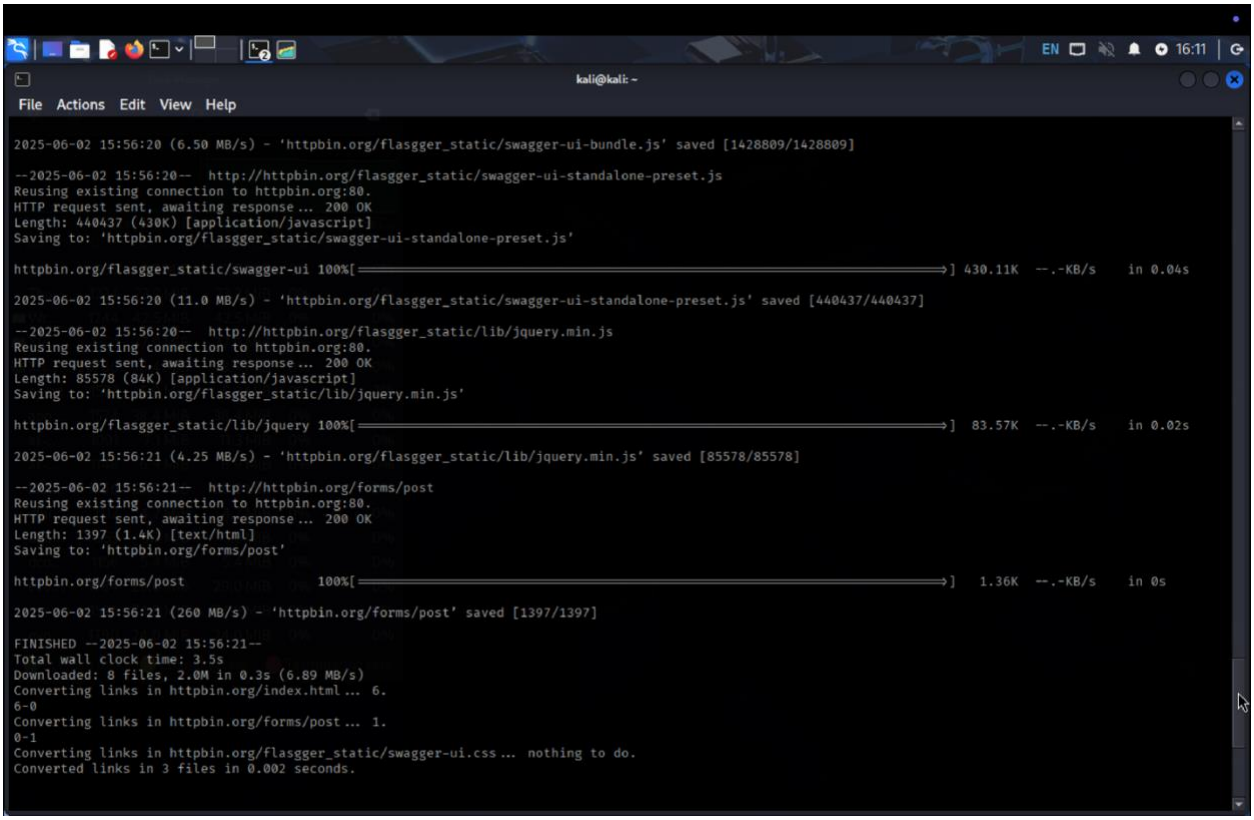

kali@kali: ~
File Actions Edit View Help
(kali㉿kali)-[~]
$ cat html
<!DOCTYPE html>
<html>
<head>
</head>
<body>
<h1>Herman Melville - Moby-Dick</h1>

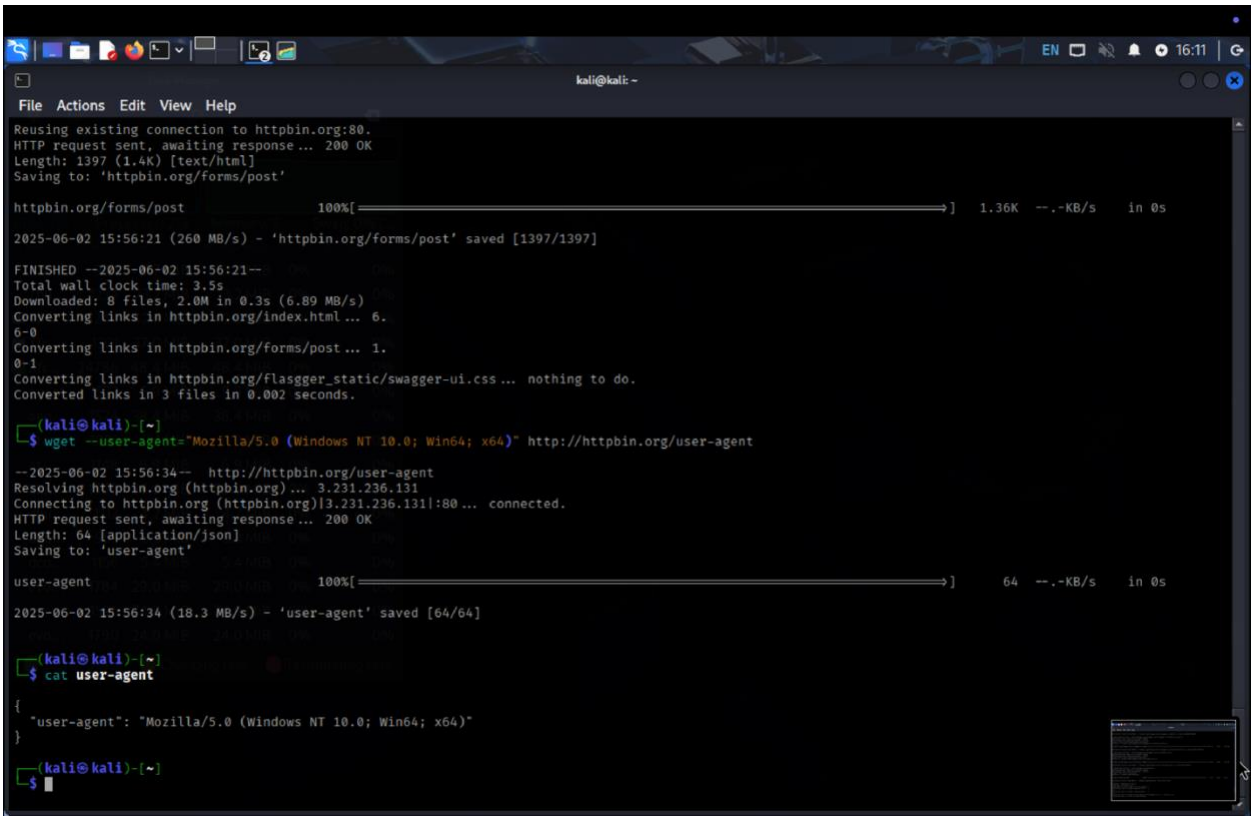
Availing himself of the mild, summer-cool weather that now reigned in these latitudes, and in preparation for the peculiarly active pursuits shortly to be anticipated, Perth, the begrimed, blistered old blacksmith, had not removed his portable forge to the hold again, after concluding his contributory work for Ahab's leg, but still retained it on deck, fast lashed to ringbolts by the foremast; being now almost incessantly invoked by the headsmen, and harpooneers, and bowsmen to do some little job for them; altering, or repairing, or new shaping their various weapons and boat furniture. Often he would be surrounded by an eager circle, all waiting to be served; holding boat-spades, pike-heads, harpoons, and lances, and jealously watching his every sooty movement, as he toiled. Nevertheless, this old man's was a patient hammer wielded by a patient arm. No murmur, no impatience, no petulance did come from him. Silent, slow, and solemn; bowing over still further his chronically broken back, he toiled away, as if toil were life itself, and the heavy beating of his hammer the heavy beating of his heart. And so it was.—Most miserable! A peculiar walk in this old man, a certain slight but painful appearing yawing in his gait, had at an early period of the voyage excited the curiosity of the mariners. And to the importunity of their persisted questionings he had finally given in; and so it came to pass that every one now knew the shameful story of his wretched fate. Belated, and not innocently, one bitter winter's midnight, on the road running between two country towns, the blacksmith half-stupidly felt the deadly numbness stealing over him, and sought refuge in a leaning, dilapidated barn. The issue was, the loss of the extremities of both feet. Out of this revelation, part by part, at last came out the four acts of the gladness, and the one long, and as yet uncatastrophied fifth act of the grief of his life's drama. He was an old man, who, at the age of nearly sixty, had postponedly encountered that thing in sorrow's technicals called ruin. He had been an artisan of famed excellence, and with plenty to do; owned a house and garden; embraced a youthful, daughter-like, loving wife, and three blithe, ruddy children; every Sunday went to a cheerful-looking church, planted in a grove. But one night, under cover of darkness, and further concealed in a most cunning disguisement, a desperate burglar slid into his happy home, and robbed them all of everything. And darker yet to tell, the blacksmith himself did ignorantly conduct this burglar into his family's heart. It was the Bottle Conjuror! Upon the opening of that fatal cork, forth flew the fiend, and shrivelled up his home. Now, for prudent, most wise, and economic reasons, the blacksmith's shop was in the basement of his dwelling, but with a separate entrance to it; so that always had the young and loving healthy wife listened with no unhappy nervousness, but with vigorous pleasure, to the stout ringing of her young-armed old husband's hammer; whose reverberations, muffled by passing through the floors and walls, came up to her, not unsweetly, in her nursery; and so, to stout Labor's iron lullaby, the blacksmith's infants were rocked to slumber. Oh, woe on woe! Oh, Death, why canst thou not sometimes be timely? Hadst thou taken this old blacksmith to thyself ere his full ruin came upon him, then had the young widow had a delicious grief, and her orphans a truly venerable, legendary sire to dream of in their after years; and all of them a care-killing competency.

</body>
</html>
(kali㉿kali)-[~]
$ wget -p -k http://httpbin.org/html


EN
16:11
kali@kali: ~
File Actions Edit View Help
(kali㉿kali)-[~]
$ wget -p -k http://httpbin.org/html
--2025-06-02 15:55:53-- http://httpbin.org/html
Resolving httpbin.org (httpbin.org)... 35.173.140.53
Connecting to httpbin.org (httpbin.org)|35.173.140.53|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 3741 (3.7K) [text/html]
Saving to: 'httpbin.org/html'
httpbin.org/html 100%[=>] 3.65K --.-KB/s in 0s
2025-06-02 15:55:55 (274 MB/s) - 'httpbin.org/html' saved [3741/3741]
FINISHED --2025-06-02 15:55:55--
Total wall clock time: 1.8s
Downloaded: 1 files, 3.7K in 0s (274 MB/s)
Converting links in httpbin.org/html... nothing to do.
Converted links in 1 files in 0 seconds.
(kali㉿kali)-[~]
$ wget -r -np -k http://httpbin.org/
--2025-06-02 15:56:18-- http://httpbin.org/
Resolving httpbin.org (httpbin.org)... 50.16.79.29
Connecting to httpbin.org (httpbin.org)|50.16.79.29|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 9593 (9.4K) [text/html]
Saving to: 'httpbin.org/index.html'
httpbin.org/index.html 100%[=>] 9.37K --.-KB/s in 0s
2025-06-02 15:56:18 (82.5 MB/s) - 'httpbin.org/index.html' saved [9593/9593]
Loading robots.txt; please ignore errors.
--2025-06-02 15:56:18-- http://httpbin.org/robots.txt
Reusing existing connection to httpbin.org:80.
HTTP request sent, awaiting response... 200 OK
Length: 30 [text/plain]
Saving to: 'httpbin.org/robots.txt'
httpbin.org/robots.txt 100%[=>] 30 --.-KB/s in 0s
Close


EN 16:11
kali@kali: ~
File Actions Edit View Help
2025-06-02 15:56:18 (82.5 MB/s) - 'httpbin.org/index.html' saved [9593/9593]
Loading robots.txt; please ignore errors.
--2025-06-02 15:56:18-- http://httpbin.org/robots.txt
Reusing existing connection to httpbin.org:80.
HTTP request sent, awaiting response... 200 OK
Length: 30 [text/plain]
Saving to: 'httpbin.org/robots.txt'
httpbin.org/robots.txt 100%[=>] 30 --.-KB/s in 0s
2025-06-02 15:56:18 (4.49 MB/s) - 'httpbin.org/robots.txt' saved [30/30]
--2025-06-02 15:56:18-- http://httpbin.org/flasgger_static/swagger-ui.css
Reusing existing connection to httpbin.org:80.
HTTP request sent, awaiting response... 200 OK
Length: 153930 (150K) [text/css]
Saving to: 'httpbin.org/flasgger_static/swagger-ui.css'
httpbin.org/flasgger_static/swagger-ui 100%[=>] 150.32K --.-KB/s in 0.03s
2025-06-02 15:56:19 (5.02 MB/s) - 'httpbin.org/flasgger_static/swagger-ui.css' saved [153930/153930]
--2025-06-02 15:56:19-- http://httpbin.org/static/favicon.ico
Reusing existing connection to httpbin.org:80.
HTTP request sent, awaiting response... 200 OK
Length: 22382 (22K) [image/vnd.microsoft.icon]
Saving to: 'httpbin.org/static/favicon.ico'
httpbin.org/static/favicon.ico 100%[=>] 21.86K --.-KB/s in 0s
2025-06-02 15:56:19 (685 MB/s) - 'httpbin.org/static/favicon.ico' saved [22382/22382]
--2025-06-02 15:56:19-- http://httpbin.org/flasgger_static/swagger-ui-bundle.js
Reusing existing connection to httpbin.org:80.
HTTP request sent, awaiting response... 200 OK
Length: 1428809 (1.4M) [application/javascript]
Saving to: 'httpbin.org/flasgger_static/swagger-ui-bundle.js'
httpbin.org/flasgger_static/swagger-ui 100%[=>] 1.36M 6.50MB/s in 0.2s
2025-06-02 15:56:20 (6.50 MB/s) - 'httpbin.org/flasgger_static/swagger-ui-bundle.js' saved [1428809/1428809]


EN 16:11
kali@kali: ~
File Actions Edit View Help
2025-06-02 15:56:20 (6.50 MB/s) - 'httpbin.org/flasgger_static/swagger-ui-bundle.js' saved [1428809/1428809]
--2025-06-02 15:56:20-- http://httpbin.org/flasgger_static/swagger-ui-standalone-preset.js
Reusing existing connection to httpbin.org:80.
HTTP request sent, awaiting response... 200 OK
Length: 440437 (430K) [application/javascript]
Saving to: 'httpbin.org/flasgger_static/swagger-ui-standalone-preset.js'
httpbin.org/flasgger_static/swagger-ui 100%[=>] 430.11K --.-KB/s in 0.04s
2025-06-02 15:56:20 (11.0 MB/s) - 'httpbin.org/flasgger_static/swagger-ui-standalone-preset.js' saved [440437/440437]
--2025-06-02 15:56:20-- http://httpbin.org/flasgger_static/lib/jquery.min.js
Reusing existing connection to httpbin.org:80.
HTTP request sent, awaiting response... 200 OK
Length: 85578 (84K) [application/javascript]
Saving to: 'httpbin.org/flasgger_static/lib/jquery.min.js'
httpbin.org/flasgger_static/lib/jquery 100%[=>] 83.57K --.-KB/s in 0.02s
2025-06-02 15:56:21 (4.25 MB/s) - 'httpbin.org/flasgger_static/lib/jquery.min.js' saved [85578/85578]
--2025-06-02 15:56:21-- http://httpbin.org/forms/post
Reusing existing connection to httpbin.org:80.
HTTP request sent, awaiting response... 200 OK
Length: 1397 (1.4K) [text/html]
Saving to: 'httpbin.org/forms/post'
httpbin.org/forms/post 100%[=>] 1.36K --.-KB/s in 0s
2025-06-02 15:56:21 (260 MB/s) - 'httpbin.org/forms/post' saved [1397/1397]
FINISHED --2025-06-02 15:56:21--
Total wall clock time: 3.5s
Downloaded: 8 files, 2.0M in 0.3s (6.89 MB/s)
Converting links in httpbin.org/index.html... 6.
6-0
Converting links in httpbin.org/forms/post... 1.
0-1
Converting links in httpbin.org/flasgger_static/swagger-ui.css... nothing to do.
Converted links in 3 files in 0.002 seconds.


kali@kali: ~
File Actions Edit View Help
Reusing existing connection to httpbin.org:80.
HTTP request sent, awaiting response... 200 OK
Length: 1397 (1.4K) [text/html]
Saving to: 'httpbin.org/forms/post'
httpbin.org/forms/post 100%[=>] 1.36K --.-KB/s in 0s
2025-06-02 15:56:21 (260 MB/s) - 'httpbin.org/forms/post' saved [1397/1397]
FINISHED --2025-06-02 15:56:21--
Total wall clock time: 3.5s
Downloaded: 8 files, 2.0M in 0.3s (6.89 MB/s)
Converting links in httpbin.org/index.html... 6.
6-0
Converting links in httpbin.org/forms/post... 1.
0-1
Converting links in httpbin.org/flasgger_static/swagger-ui.css... nothing to do.
Converted links in 3 files in 0.002 seconds.
(kali㉿kali)-[~]
$ wget --user-agent="Mozilla/5.0 (Windows NT 10.0; Win64; x64)" http://httpbin.org/user-agent
--2025-06-02 15:56:34-- http://httpbin.org/user-agent
Resolving httpbin.org (httpbin.org)... 3.231.236.131
Connecting to httpbin.org (httpbin.org)|3.231.236.131|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 64 [application/json]
Saving to: 'user-agent'
user-agent 100%[=>] 64 --.-KB/s in 0s
2025-06-02 15:56:34 (18.3 MB/s) - 'user-agent' saved [64/64]
(kali㉿kali)-[~]
$ cat user-agent
{
"user-agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64)"
}
(kali㉿kali)-[~]
$