iGNITE
Technologies

A Detailed Guide on Cewl

WWW.HACKINGARTICLES.IN

Contents

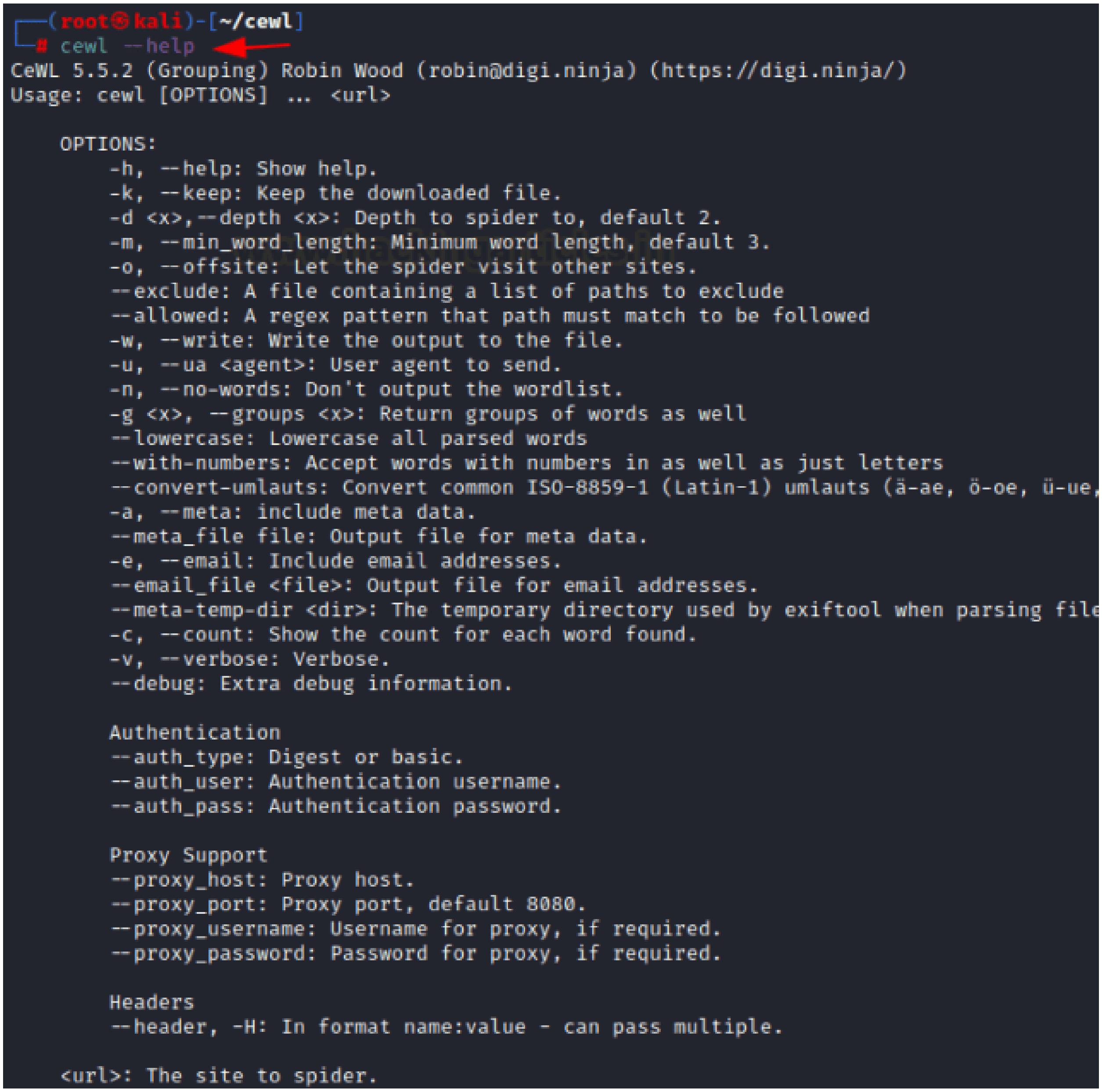
Introduction

CeWL – A custom wordlist generator is a ruby program that crawls a specific URL to a defined depth and returns a list of keywords, which password crackers like John the Ripper, Medusa, and WFuzz can use to crack the passwords. Cewl also has an associated command-line app FAB, which uses the same metadata extraction techniques to generate author/producer lists from already downloaded files using information extraction algorithms like CeWL.

CeWL comes preinstalled with Kali Linux. With this tool, we can easily collect words and phrases from the target page. It is a robust program that can quickly scrape the webserver of any website.

Open the terminal of Kali Linux and type "cewl -h" to see the lists of all the options it accepts, with a complete description.

Syntax: cewl <url> [options]



```
┌──(root㉿kali)-[~/cewl]
└─# cewl --help
CeWL 5.5.2 (Grouping) Robin Wood (robin@digi.ninja) (https://digi.ninja/)
Usage: cewl [OPTIONS] ... <url>

    OPTIONS:
        -h, --help: Show help.
        -k, --keep: Keep the downloaded file.
        -d <x>,--depth <x>: Depth to spider to, default 2.
        -m, --min_word_length: Minimum word length, default 3.
        -o, --offsite: Let the spider visit other sites.
        --exclude: A file containing a list of paths to exclude
        --allowed: A regex pattern that path must match to be followed
        -w, --write: Write the output to the file.
        -u, --ua <agent>: User agent to send.
        -n, --no-words: Don't output the wordlist.
        -g <x>, --groups <x>: Return groups of words as well
        --lowercase: Lowercase all parsed words
        --with-numbers: Accept words with numbers in as well as just letters
        --convert-umlauts: Convert common ISO-8859-1 (Latin-1) umlauts (ä-ae, ö-oe, ü-ue,
        -a, --meta: include meta data.
        --meta_file file: Output file for meta data.
        -e, --email: Include email addresses.
        --email_file <file>: Output file for email addresses.
        --meta-temp-dir <dir>: The temporary directory used by exiftool when parsing file
        -c, --count: Show the count for each word found.
        -v, --verbose: Verbose.
        --debug: Extra debug information.

        Authentication
        --auth_type: Digest or basic.
        --auth_user: Authentication username.
        --auth_pass: Authentication password.

        Proxy Support
        --proxy_host: Proxy host.
        --proxy_port: Proxy port, default 8080.
        --proxy_username: Username for proxy, if required.
        --proxy_password: Password for proxy, if required.

        Headers
        --header, -H: In format name:value - can pass multiple.

    <url>: The site to spider.
```

General Options :

-h, –help:	Show help.
-k, –keep:	Keep the downloaded file.
-d <x>, –depth <x>:	Depth to spider to, default 2.
-m, –min_word_length:	Minimum word length, default 3.
-o, –offsite:	Let the spider visit other sites.
-w, –write:	Write the output to the file.
-u, –ua <agent>:	User agent to send.

-n, –no-words:	Don’t output the wordlist.
–with-numbers:	Accept words with numbers in as well as just letters
-a, –meta:	include meta data.
–meta_file file:	Output file for Meta data.
-e, –email:	Include email addresses.
–email_file <file>:	Output file for email addresses.
-c, –count:	Show the count for each word found.
-v, –verbose:	Verbose.
–debug:	Extra debug information

Authentication

–auth_type:	Digest or basic.
–auth_user:	Authentication username.
–auth_pass:	Authentication password.

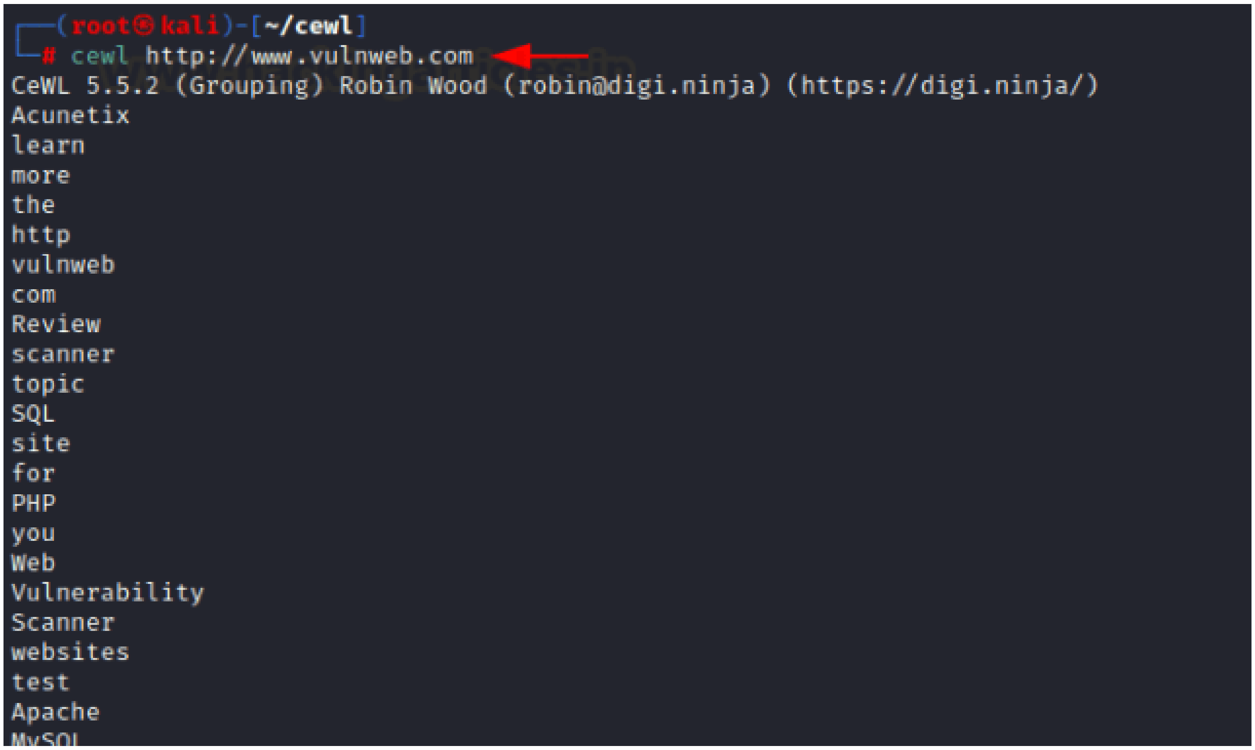
Proxy Support

–proxy_host:	Proxy host.
–proxy_port:	Proxy port, default 8080.
–proxy_username:	Username for proxy, if required.
–proxy_password:	Password for proxy, if required.

Default Procedure

Use the following command to generate a list of words that will spider the given URL to a specified depth and we can use it as a directory for cracking the passwords.

```
cewl http://www.vulnweb.com
```



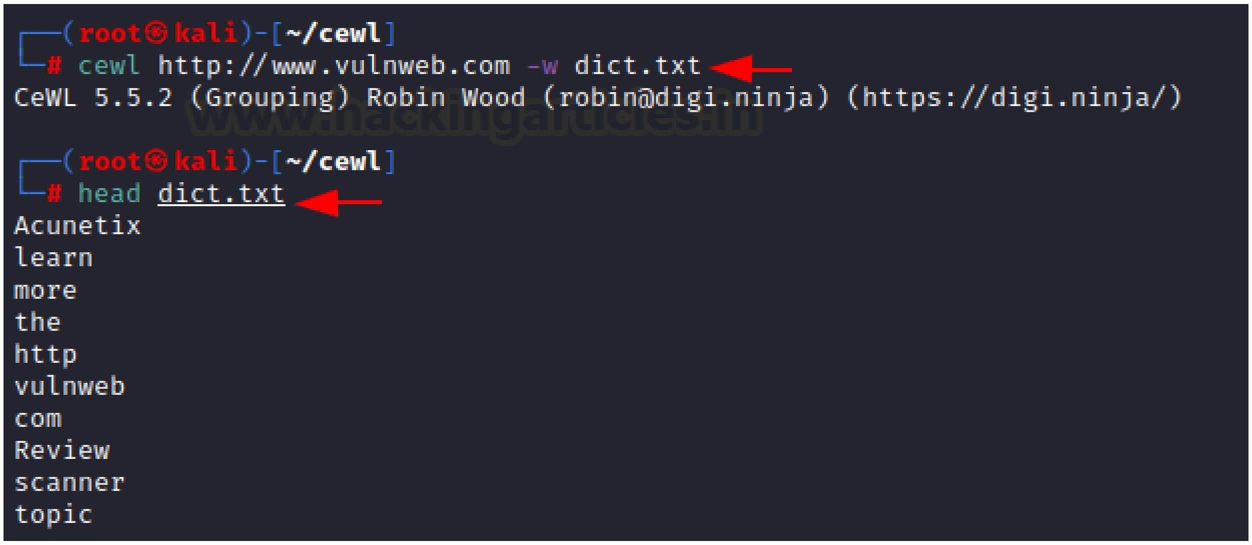
```
┌──(root㉿kali)-[~/cewl]
└─# cewl http://www.vulnweb.com
CeWL 5.5.2 (Grouping) Robin Wood (robin@digi.ninja) (https://digi.ninja/)
Acunetix
learn
more
the
http
vulnweb
com
Review
scanner
topic
SQL
site
for
PHP
you
Web
Vulnerability
Scanner
websites
test
Apache
MySQL
```

Store this wordlist in a file

Now to save this all wordlist in a file for record-keeping, efficiency and readability we will use the -w option to save the output in a text file.

```
cewl http://www.vulnweb.com -w dict.txt
```

Here dict.txt is the file name where the wordlist will be stored. Once the file has been created you can open it to see if the output is stored in the file.

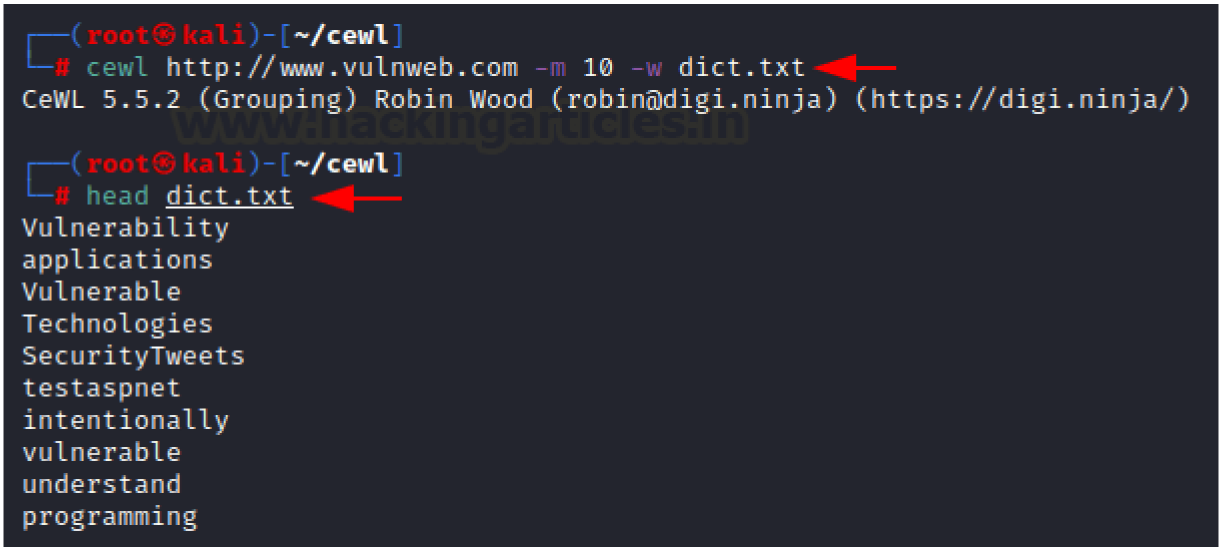


```
┌──(root㉿kali)-[~/cewl]
└─# cewl http://www.vulnweb.com -w dict.txt
CeWL 5.5.2 (Grouping) Robin Wood (robin@digi.ninja) (https://digi.ninja/)

┌──(root㉿kali)-[~/cewl]
└─# head dict.txt
Acunetix
learn
more
the
http
vulnweb
com
Review
scanner
topic
```

Generating wordlists of a certain length

If you want to create a wordlist of a specific length then you can choose to use option -m and provide the minimum length for the keyword hence it will create wordlists for a certain length.

```
cewl http://vulnweb.com / -m 10 -w dict.txt
```



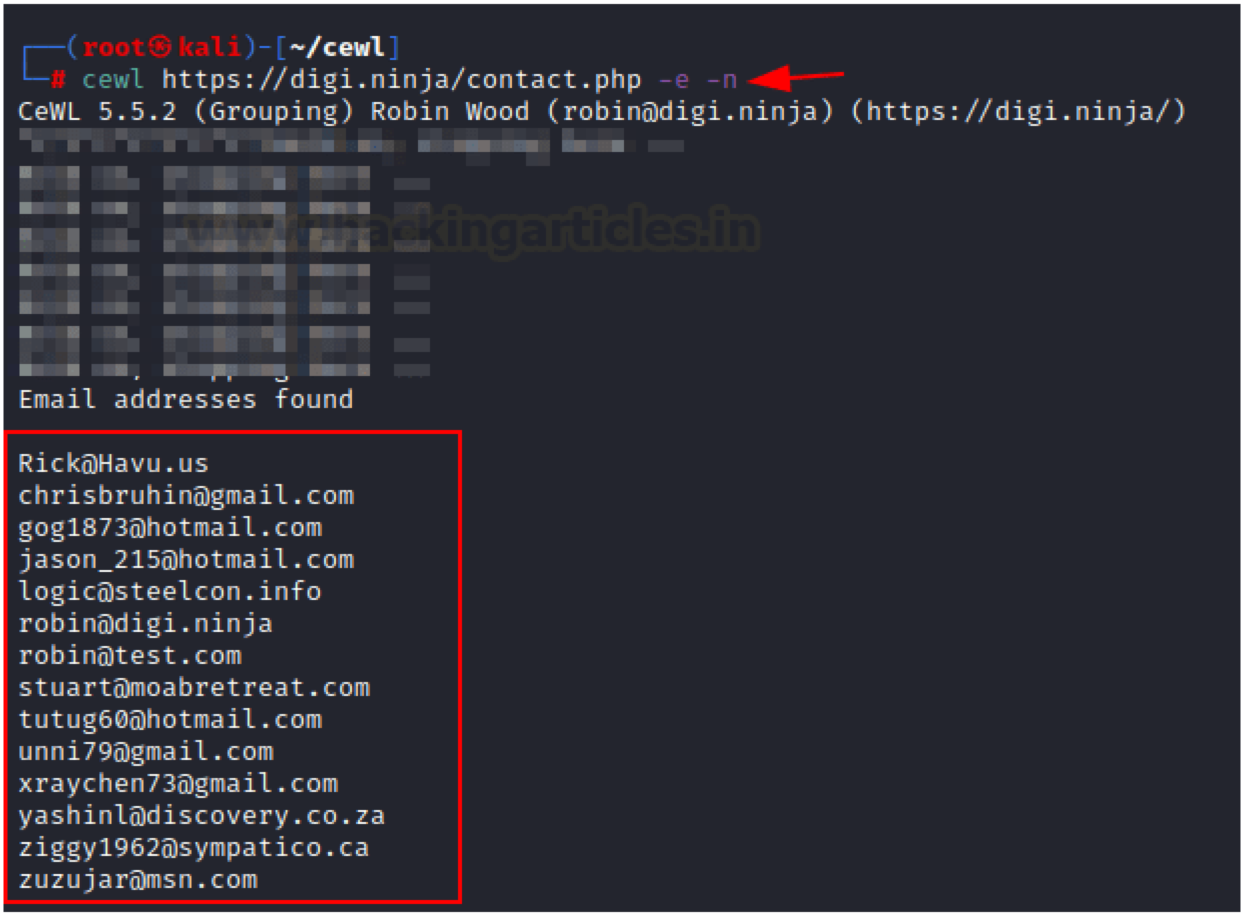
```
┌──(root㉿kali)-[~/cewl]
└─# cewl http://www.vulnweb.com -m 10 -w dict.txt
CeWL 5.5.2 (Grouping) Robin Wood (robin@digi.ninja) (https://digi.ninja/)

┌──(root㉿kali)-[~/cewl]
└─# head dict.txt
Vulnerability
applications
Vulnerable
Technologies
SecurityTweets
testaspnet
intentionally
vulnerable
understand
programming
```

So basically, this will create a wordlist in which each word has a minimum of 10 letters and store these keywords in the file dict.txt. Screenshot is attached for your reference.

Retrieval of Emails from the website:

In order to retrieve emails from the website, we can use the -e option, while the -n option will hide the lists created while crawling the provided website. As you can see in the screenshot attached it has found 1 email-id from the website.

```
cewl https://digi.ninja/contact.php -e -n
```



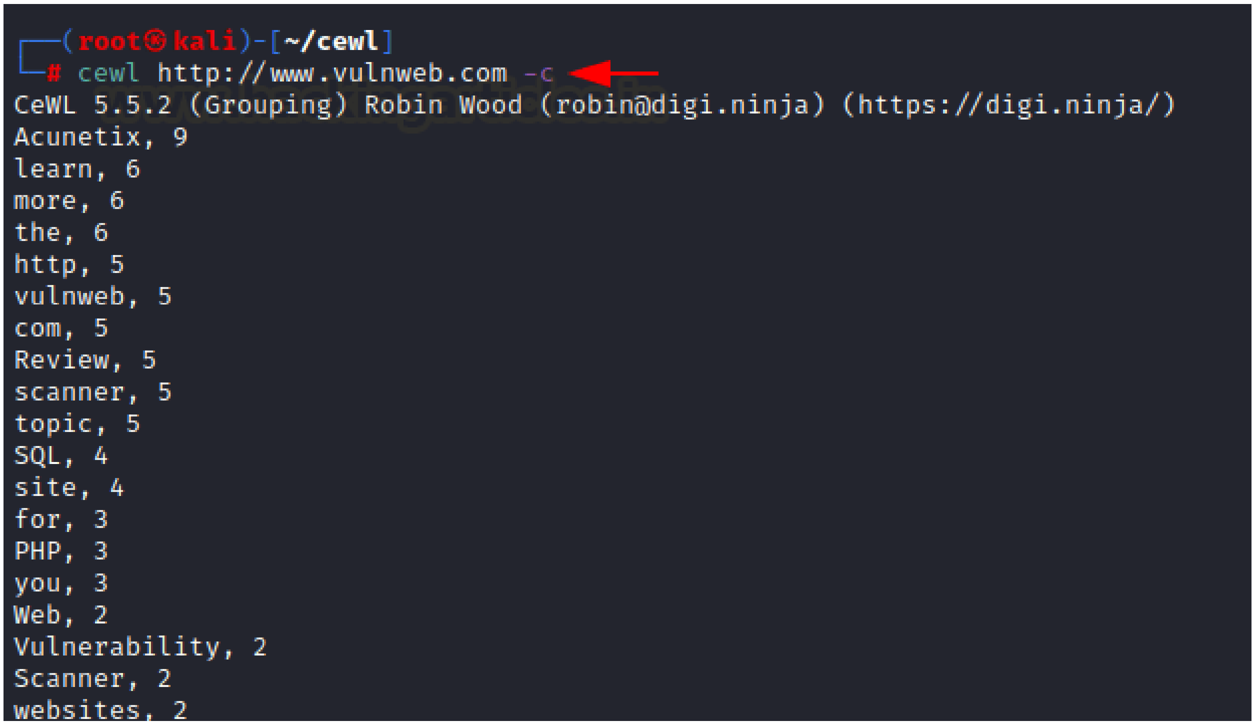
```
┌──(root㉿kali)-[~/cewl]
└─# cewl https://digi.ninja/contact.php -e -n
CeWL 5.5.2 (Grouping) Robin Wood (robin@digi.ninja) (https://digi.ninja/)
[redacted]
Email addresses found

Rick@Havu.us
chrisbruhin@gmail.com
gog1873@hotmail.com
jason_215@hotmail.com
logic@steelcon.info
robin@digi.ninja
robin@test.com
stuart@moabretreat.com
tutug60@hotmail.com
unni79@gmail.com
xraychen73@gmail.com
yashinl@discovery.co.za
ziggy1962@sympatico.ca
zuzujar@msn.com
```

To count the number of words repeated on the website

If you want to count the number of times a word is repeated on a website, then use the -c option that will enable the count parameter.

```
cewl http://www.vulnweb.com -c
```

For your reference, a screenshot is added below which prints the count for every keyword repeated on website.

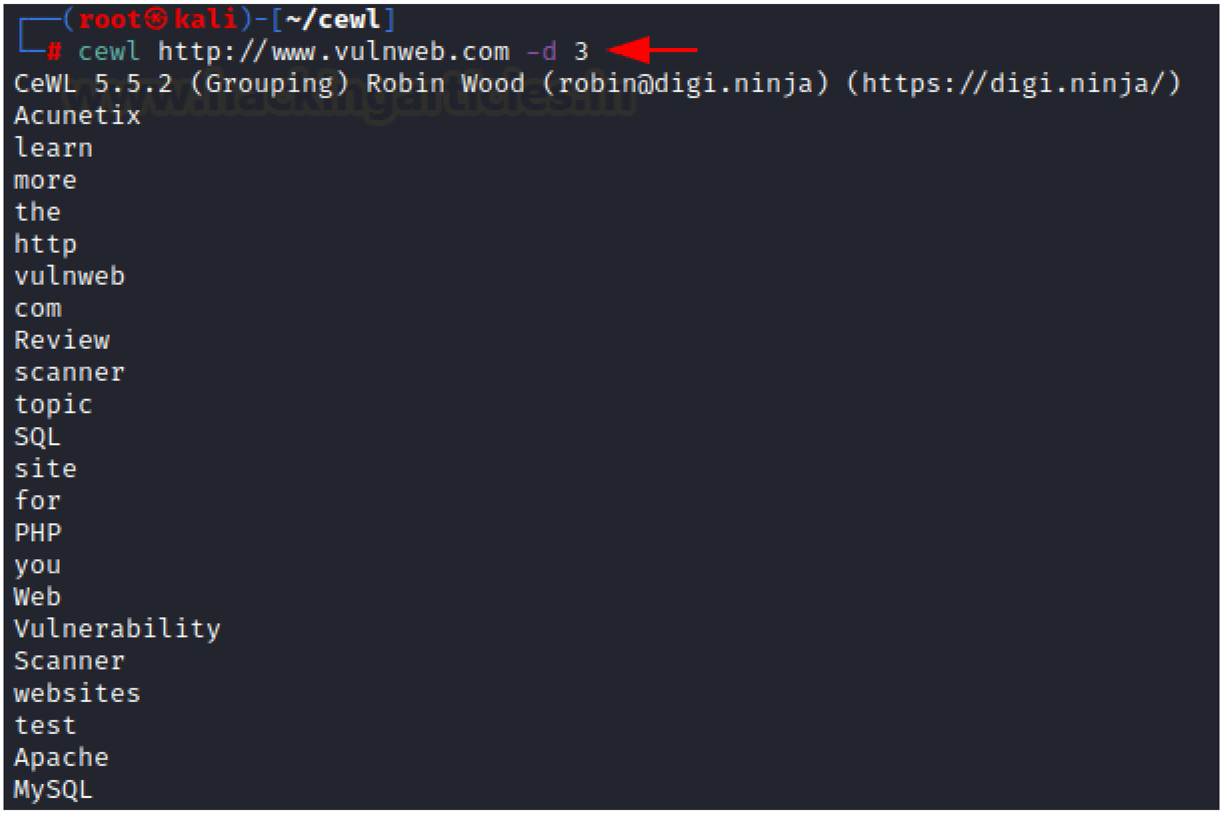


```
┌──(root㉿kali)-[~/cewl]
└─# cewl http://www.vulnweb.com -c
CeWL 5.5.2 (Grouping) Robin Wood (robin@digi.ninja) (https://digi.ninja/)
Acunetix, 9
learn, 6
more, 6
the, 6
http, 5
vulnweb, 5
com, 5
Review, 5
scanner, 5
topic, 5
SQL, 4
site, 4
for, 3
PHP, 3
you, 3
Web, 2
Vulnerability, 2
Scanner, 2
websites, 2
```

Increase Spider depth

You can use -d option with the depth number to activate depth parameter for more quick and intense crawling so that a large list of words is created. The depth level is set to 2 as default.

```
cewl http://vulnweb.com -d 3
```



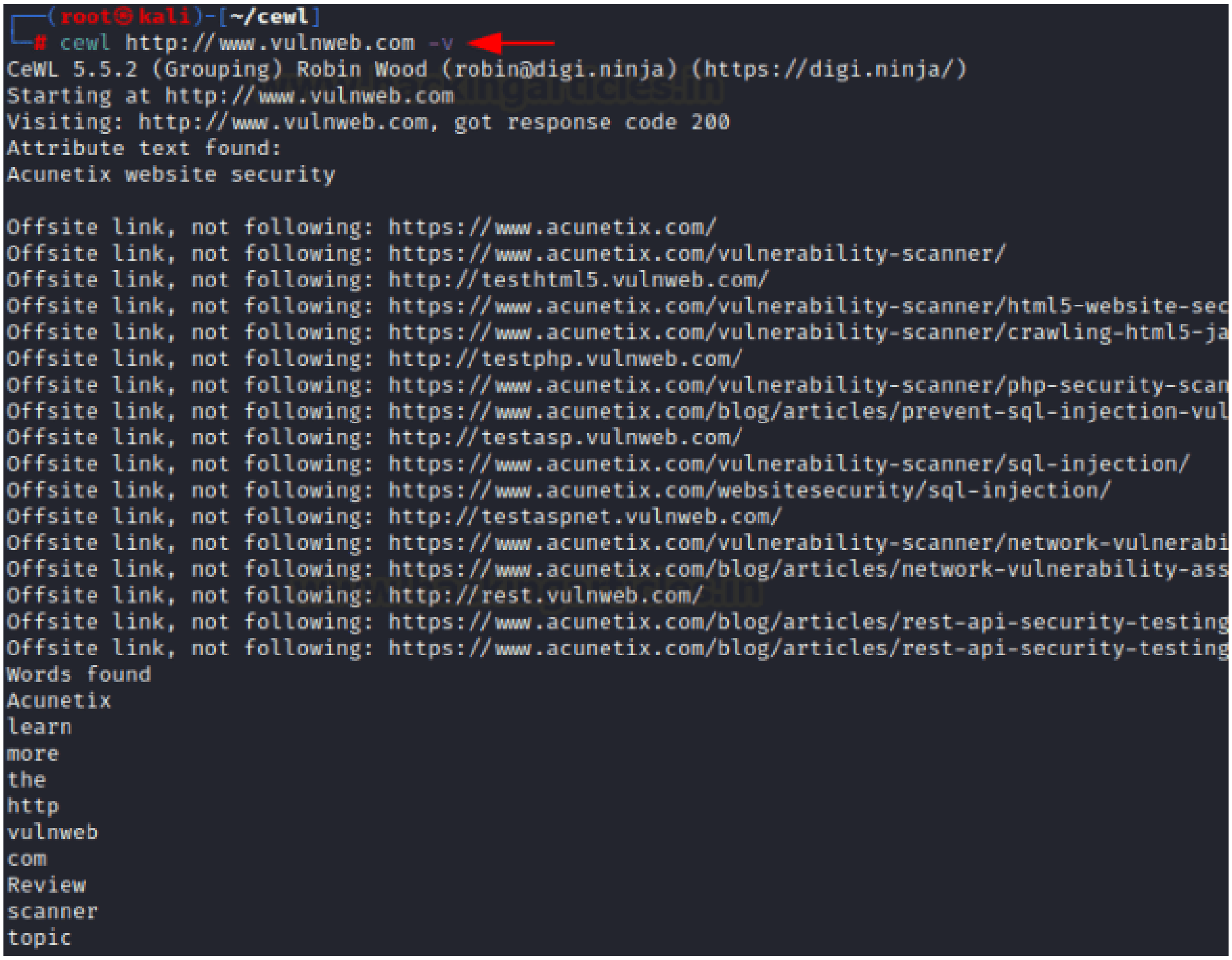
```
┌──(root㉿kali)-[~/cewl]
└─# cewl http://www.vulnweb.com -d 3
CeWL 5.5.2 (Grouping) Robin Wood (robin@digi.ninja) (https://digi.ninja/)
Acunetix
learn
more
the
http
vulnweb
com
Review
scanner
topic
SQL
site
for
PHP
you
Web
Vulnerability
Scanner
websites
test
Apache
MySQL
```

Verbose Mode

We have a -v option for the verbose mode to extend the website crawling result and retrieve complete detail of the website.

```
cewl http://vulnweb.com -v
```

So, this will display extended website crawling results. Below we have attached a screenshot so that you will get a clear idea.

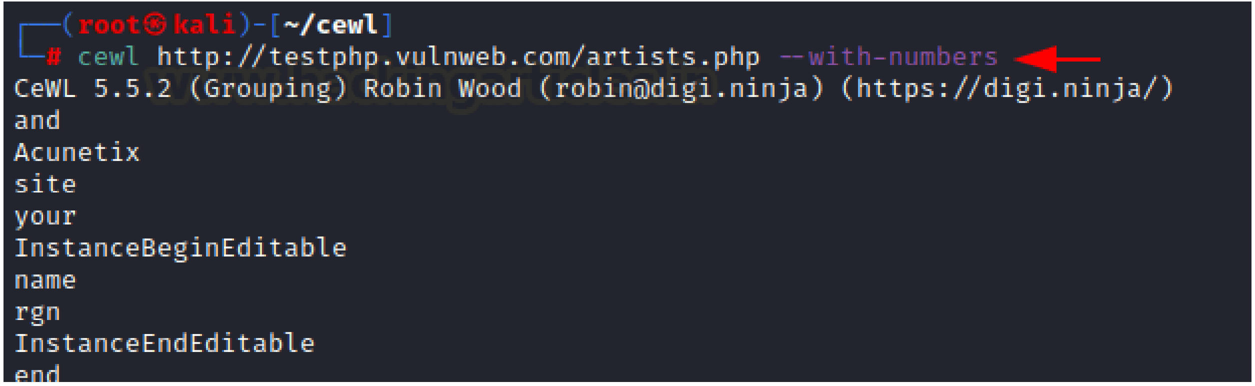


```
┌──(root㉿kali)-[~/cewl]
└─# cewl http://www.vulnweb.com -v
CeWL 5.5.2 (Grouping) Robin Wood (robin@digi.ninja) (https://digi.ninja/)
Starting at http://www.vulnweb.com
Visiting: http://www.vulnweb.com, got response code 200
Attribute text found:
Acunetix website security

Offsite link, not following: https://www.acunetix.com/
Offsite link, not following: https://www.acunetix.com/vulnerability-scanner/
Offsite link, not following: http://testhtml5.vulnweb.com/
Offsite link, not following: https://www.acunetix.com/vulnerability-scanner/html5-website-sec
Offsite link, not following: https://www.acunetix.com/vulnerability-scanner/crawling-html5-ja
Offsite link, not following: http://testphp.vulnweb.com/
Offsite link, not following: https://www.acunetix.com/vulnerability-scanner/php-security-scan
Offsite link, not following: https://www.acunetix.com/blog/articles/prevent-sql-injection-vul
Offsite link, not following: http://testasp.vulnweb.com/
Offsite link, not following: https://www.acunetix.com/vulnerability-scanner/sql-injection/
Offsite link, not following: https://www.acunetix.com/websitesecurity/sql-injection/
Offsite link, not following: http://testaspnet.vulnweb.com/
Offsite link, not following: https://www.acunetix.com/vulnerability-scanner/network-vulnerabi
Offsite link, not following: https://www.acunetix.com/blog/articles/network-vulnerability-ass
Offsite link, not following: http://rest.vulnweb.com/
Offsite link, not following: https://www.acunetix.com/blog/articles/rest-api-security-testing
Offsite link, not following: https://www.acunetix.com/blog/articles/rest-api-security-testing
Words found
Acunetix
learn
more
the
http
vulnweb
com
Review
scanner
topic
```

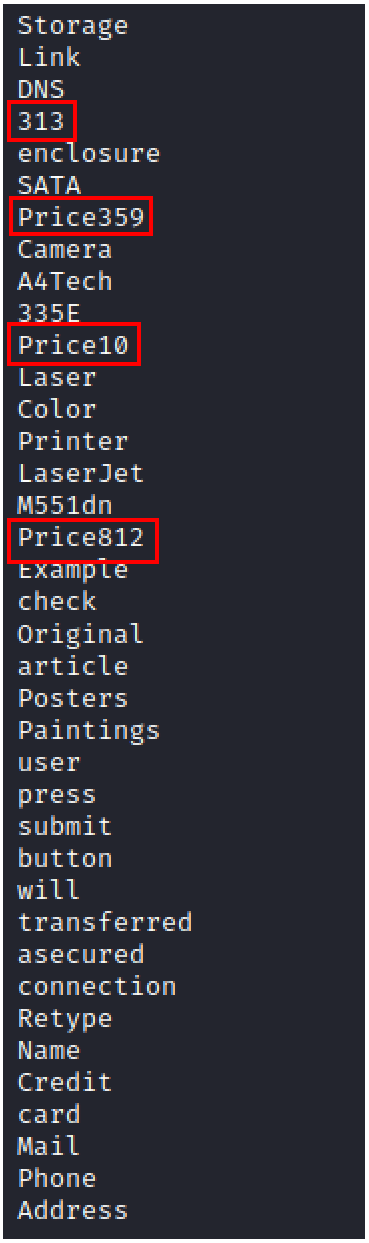
Alphanumeric Wordlist

Sometimes it may happen that you may need an alpha-numeric wordlist that you can use –the with-numbers option to get an alpha-numeric wordlist.

```
cewl http://testphp.vulnweb.com/artists.php --with-numbers
```



```
┌──(root㉿kali)-[~/cewl]
└─# cewl http://testphp.vulnweb.com/artists.php --with-numbers
CeWL 5.5.2 (Grouping) Robin Wood (robin@digi.ninja) (https://digi.ninja/)
and
Acunetix
site
your
InstanceBeginEditable
name
rgn
InstanceEndEditable
end
```



```
Storage
Link
DNS
313
enclosure
SATA
Price359
Camera
A4Tech
335E
Price10
Laser
Color
Printer
LaserJet
M551dn
Price812
Example
check
Original
article
Posters
Paintings
user
press
submit
button
will
transferred
asecured
connection
Retype
Name
Credit
card
Mail
Phone
Address
```

Cewl with Digest/Basic Authentication

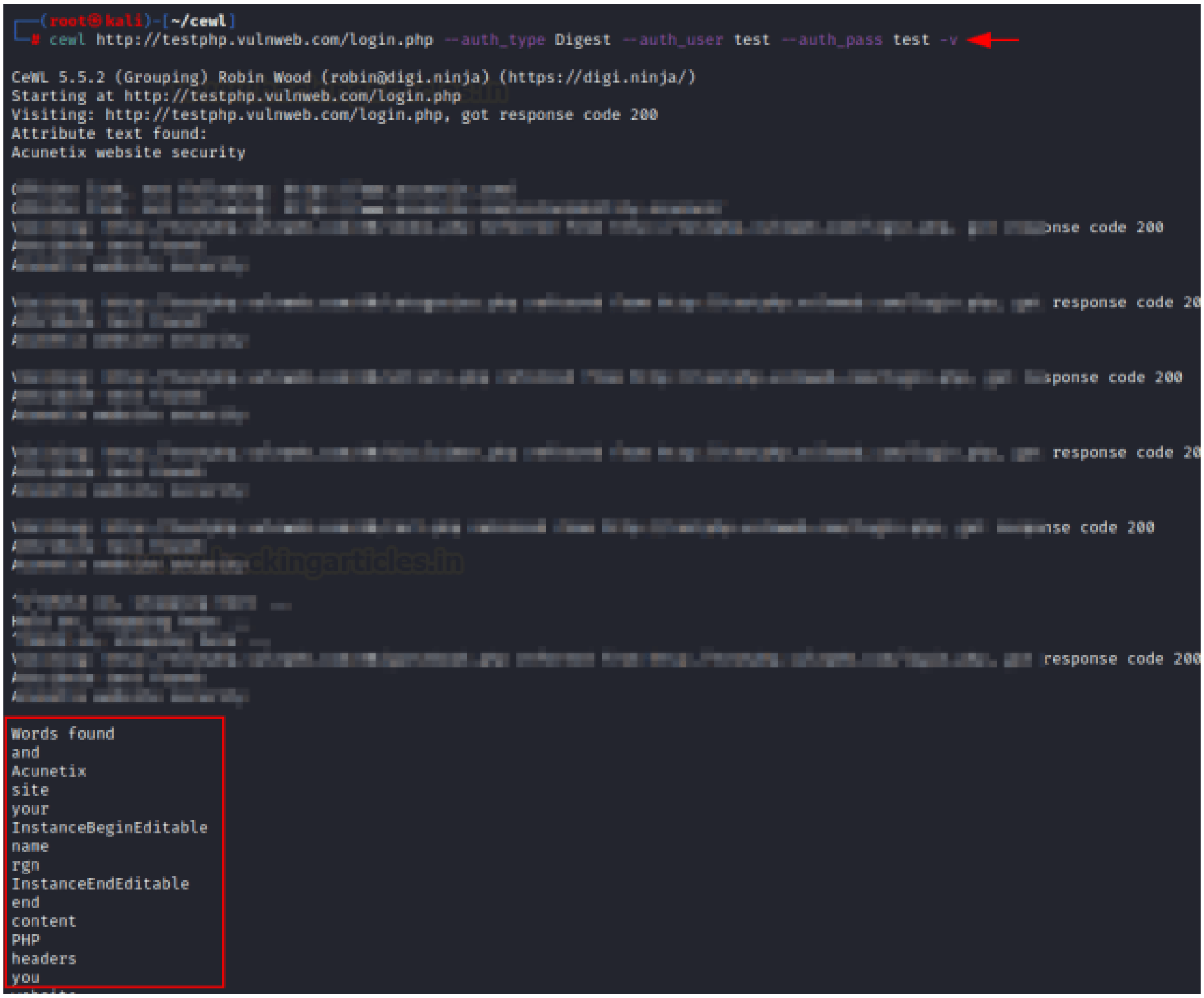
It may happen sometimes that some web applications may have an authentication page for login and for that the above basic command will not give desired results. So for that, you need to bypass the authentication page by using the command given below.

```
cewl http://testphp.vulnweb.com/login.php --auth_type Digest --auth_user test –auth_pass test -v
```

In this command we have used the following options:

–auth_type:	Digest /Basic
–auth_user:	Authentication Username

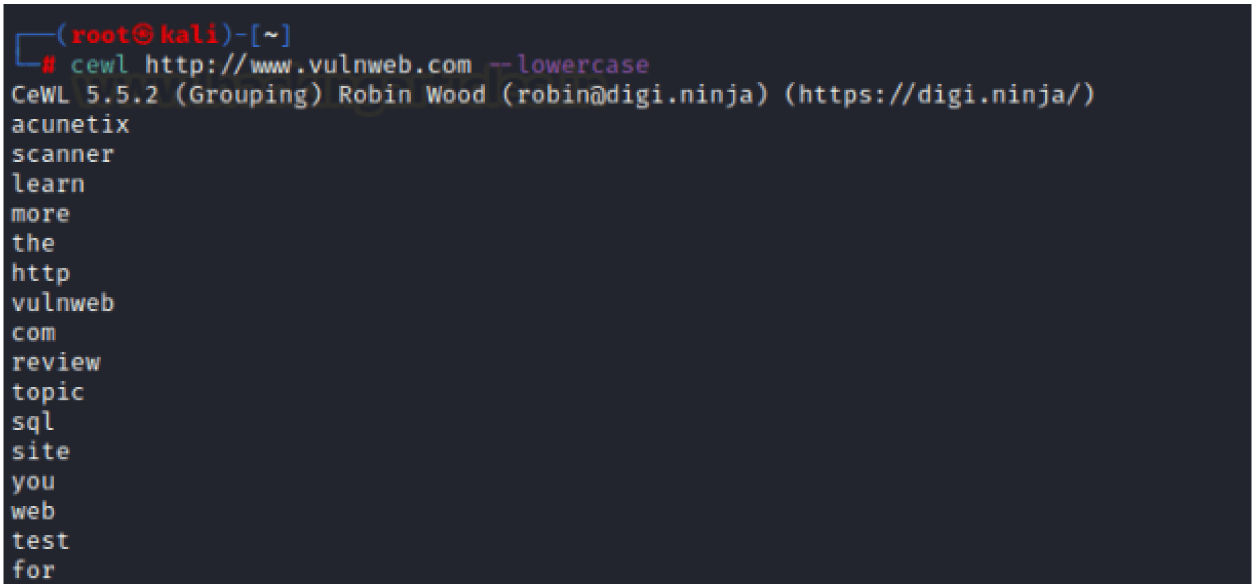
–auth_pass: Authentication password



```
┌──(root㉿kali)-[~/cewl]
└─# cewl http://testphp.vulnweb.com/login.php --auth_type Digest --auth_user test --auth_pass test -v

CeWL 5.5.2 (Grouping) Robin Wood (robin@digi.ninja) (https://digi.ninja/)
Starting at http://testphp.vulnweb.com/login.php
Visiting: http://testphp.vulnweb.com/login.php, got response code 200
Attribute text found:
Acunetix website security

[illegible]
[illegible]
[illegible]onse code 200
[illegible]
[illegible]

[illegible] response code 20
[illegible]
[illegible]

[illegible]sponse code 200
[illegible]
[illegible]

[illegible] response code 20
[illegible]
[illegible]

[illegible]nse code 200
[illegible]
[illegible]

[illegible]
[illegible]
[illegible]
[illegible] response code 200
[illegible]
[illegible]

Words found
and
Acunetix
site
your
InstanceBeginEditable
name
rgn
InstanceEndEditable
end
content
PHP
headers
you
```

Lowercase all parsed words

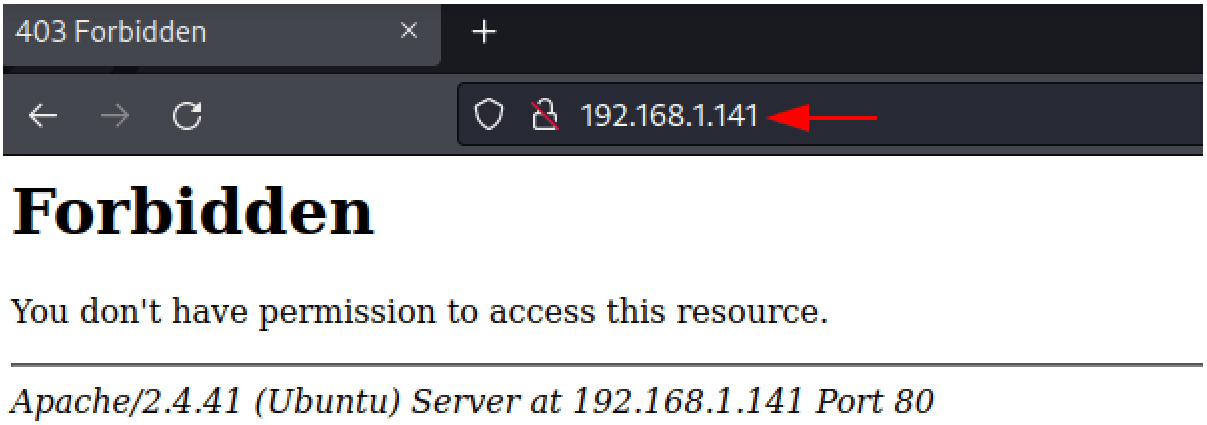
When you need the keywords to be generated in lowercase for that you can use the –lowercase option to generate the words in lowercase.



```
┌──(root㉿kali)-[~]
└─# cewl http://www.vulnweb.com --lowercase
CeWL 5.5.2 (Grouping) Robin Wood (robin@digi.ninja) (https://digi.ninja/)
acunetix
scanner
learn
more
the
http
vulnweb
com
review
topic
sql
site
you
web
test
for
```

Proxy Support

This default command for cewl will not work properly if you have attached a proxy server. We tried to access the application through ip address but the proxy server is attached hence this gave us a Forbidden Error page.



Forbidden

You don't have permission to access this resource.

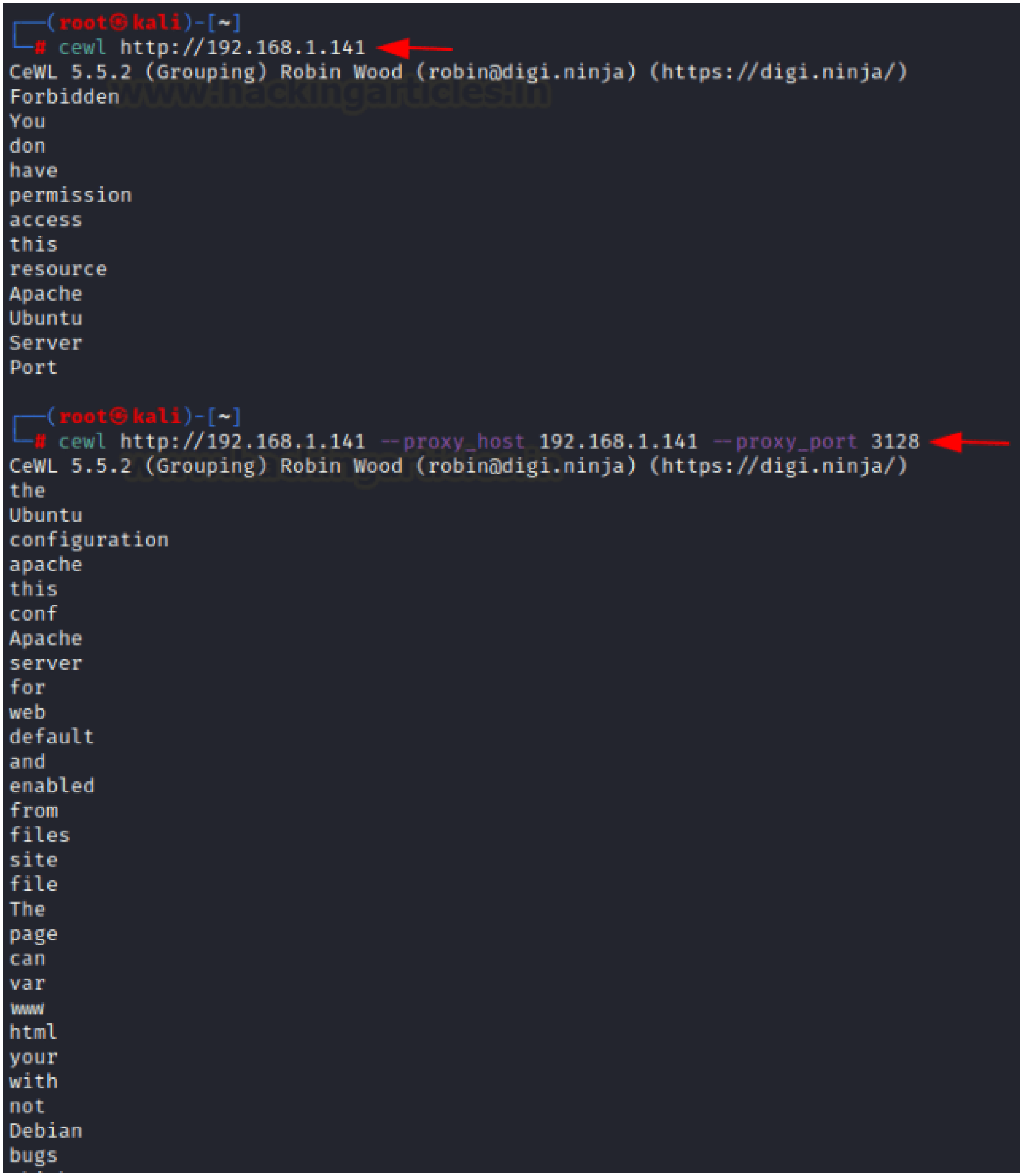
Apache/2.4.41 (Ubuntu) Server at 192.168.1.141 Port 80

And here if we apply the default cewl command so it will generate the error page wordlist. Hence to get the appropriate wordlist of the web application we have used commands as:

```
cewl http://192.168.1.141 --proxy_host 192.168.1.141 --proxy_port 3128
```

In this command we have used the following options:

–proxy_host:	Your Host
–proxy_port:	Port number of your proxy

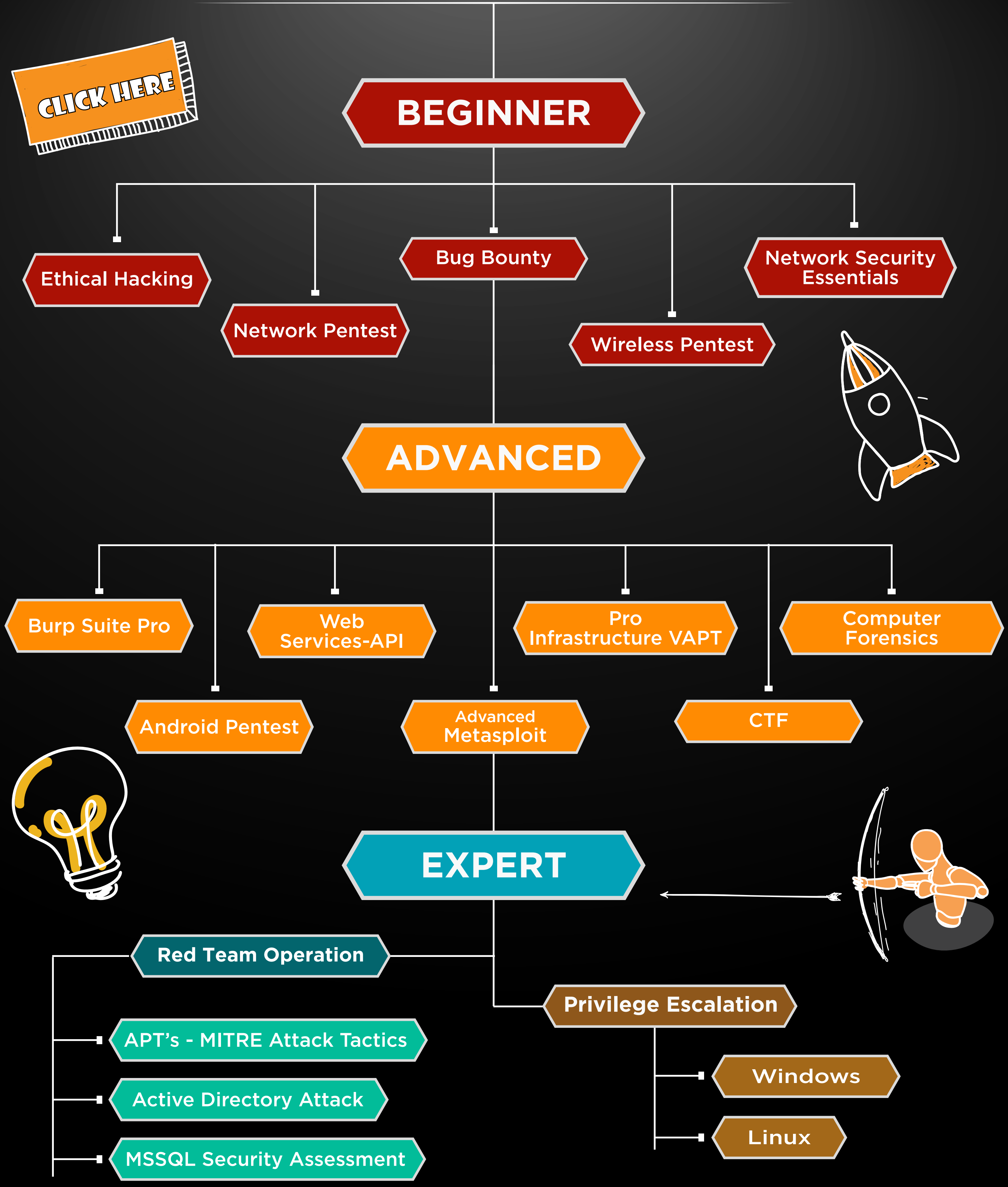


```
┌──(root㉿kali)-[~]
└─# cewl http://192.168.1.141
CeWL 5.5.2 (Grouping) Robin Wood (robin@digi.ninja) (https://digi.ninja/)
Forbidden
You
don
have
permission
access
this
resource
Apache
Ubuntu
Server
Port

┌──(root㉿kali)-[~]
└─# cewl http://192.168.1.141 --proxy_host 192.168.1.141 --proxy_port 3128
CeWL 5.5.2 (Grouping) Robin Wood (robin@digi.ninja) (https://digi.ninja/)
the
Ubuntu
configuration
apache
this
conf
Apache
server
for
web
default
and
enabled
from
files
site
file
The
page
can
var
www
html
your
with
not
Debian
bugs
```

iGNITE
Technologies

JOIN OUR TRAINING PROGRAMS

CLICK HERE



BEGINNER

- Ethical Hacking
- Network Pentest
- Bug Bounty
- Wireless Pentest
- Network Security Essentials

ADVANCED

- Burp Suite Pro
- Android Pentest
- Web Services-API
- Advanced Metasploit
- Pro Infrastructure VAPT
- CTF
- Computer Forensics

EXPERT

- Red Team Operation
  - APT's - MITRE Attack Tactics
  - Active Directory Attack
  - MSSQL Security Assessment
- Privilege Escalation
  - Windows
  - Linux

www.ignitetechnologies.in