

Content Discovery Tryhackme walkthrough



mrwhite18

Follow

4 min read · Sep 14, 2024



Jr pentester path

In this blog, I'll walk you through my experience solving the “**Content Discovery**” room on TryHackMe. This challenge is all about finding hidden pages and directories that aren't immediately visible, which can be a crucial part of any penetration testing process

Use the link to access the lab:

<https://tryhackme.com/r/room/contentdiscovery>

Task:1

What Is Content Discovery?

In web security, **content** refers to anything on a website — files, videos, images, backups, or hidden features. **Content discovery** is about finding the

things not immediately visible, often not meant for public access. This could include admin panels, backup files, older site versions, or staff-only pages.

Manually, Automated and OSINT (Open-Source Intelligence).

Start the AttackBox (by clicking the blue "Start AttackBox" button), and the machine on this task.

Answer the questions below

What is the Content Discovery method that begins with M?

Manually ✓ Correct Answer

What is the Content Discovery method that begins with A?

Automated ✓ Correct Answer

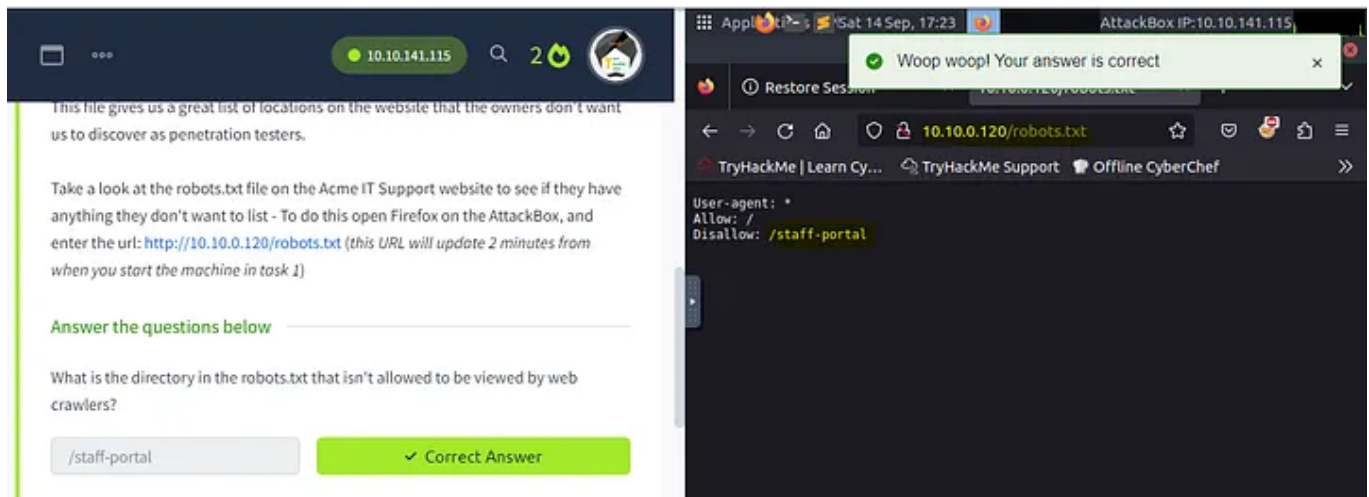
What is the Content Discovery method that begins with O?

OSINT ✓ Correct Answer

Content discovery

Task:2

Manual Discovery — Robots.txt

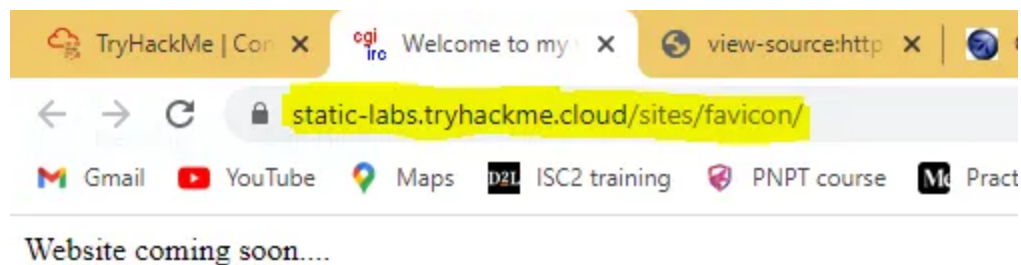


Task:3

Manual Discovery — Favicon

Favicon

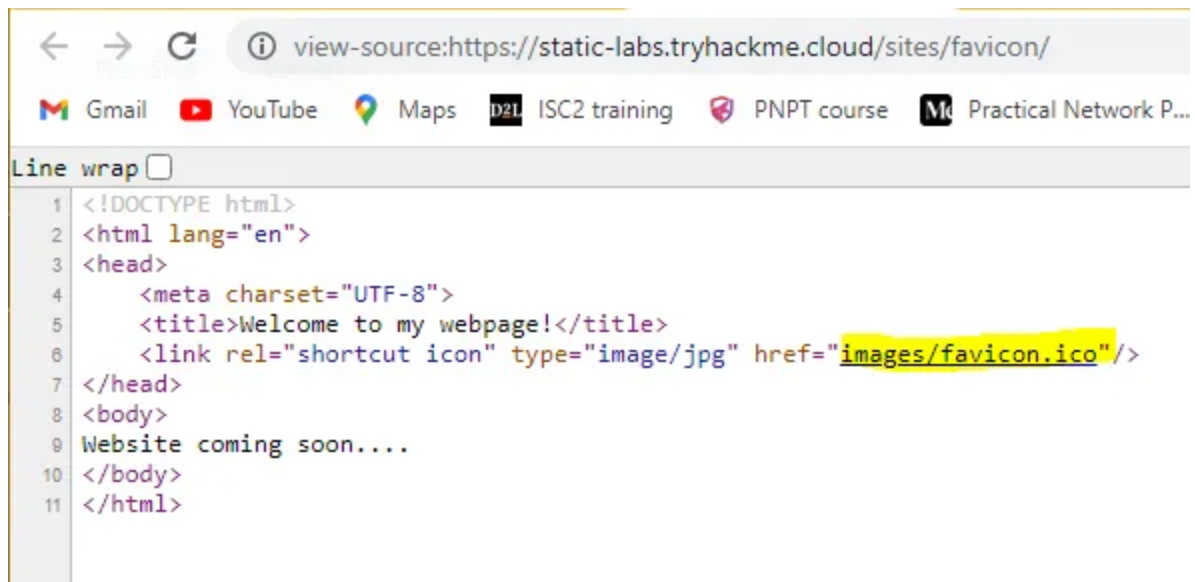
The favicon is a small icon displayed in the browser's address bar or tab used for branding a website.



Sometimes when frameworks are used to build a website, a favicon that is part of the installation gets leftover, and if the website developer doesn't replace this with a custom one, this can give us a clue on what framework is in use. OWASP host a database of common framework icons that you can use

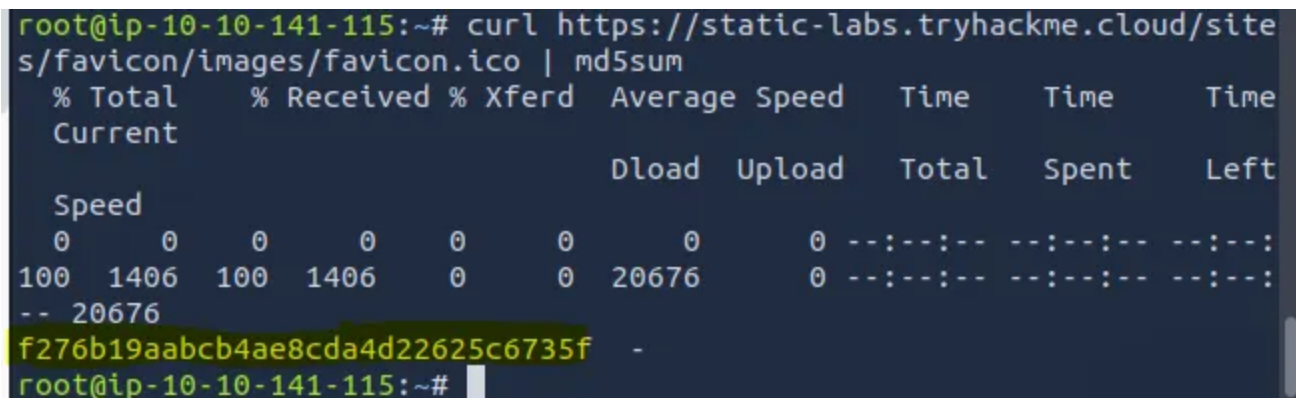
to check against the targets favicon

https://wiki.owasp.org/index.php/OWASP_favicon_database



```
1 <!DOCTYPE html>
2 <html lang="en">
3 <head>
4     <meta charset="UTF-8">
5     <title>Welcome to my webpage!</title>
6     <link rel="shortcut icon" type="image/jpeg" href="images/favicon.ico"/>
7 </head>
8 <body>
9 Website coming soon....
10 </body>
11 </html>
```

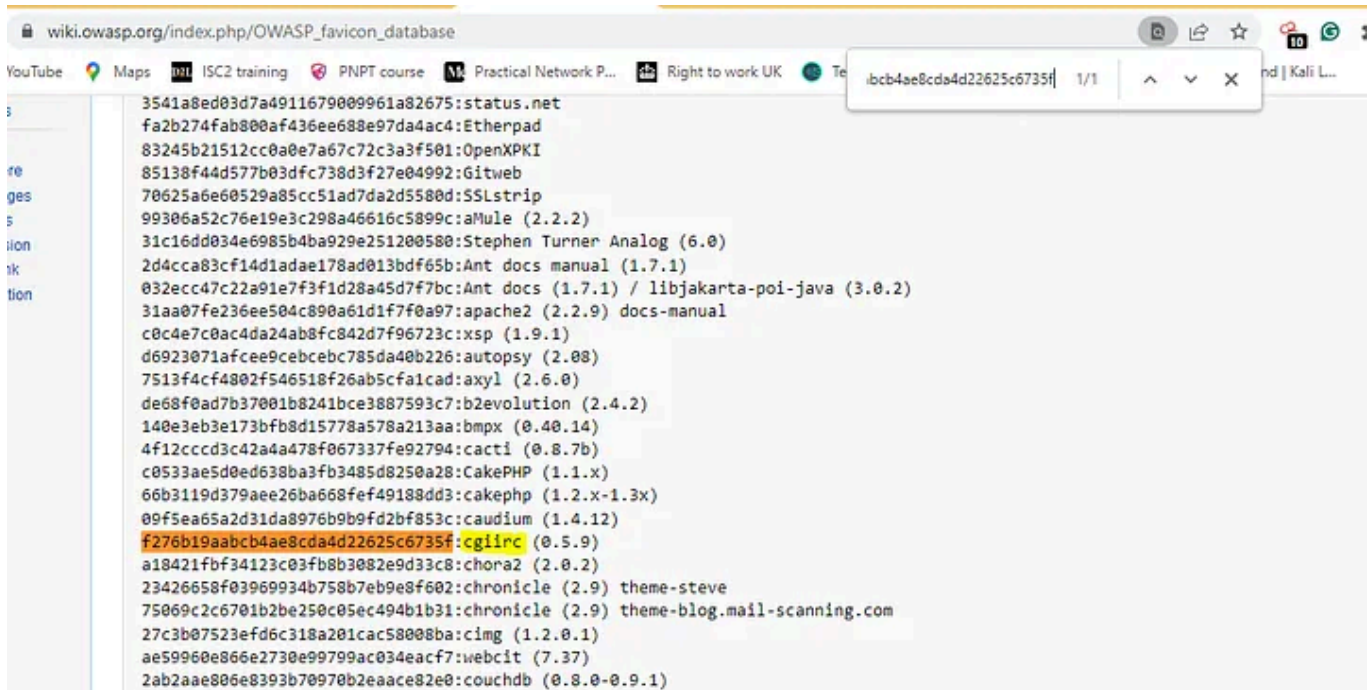
Right Click then click to view page source. then copy the link address on href =..... then paste it on the terminal as follows below screenshot



```
root@ip-10-10-141-115:~# curl https://static-labs.tryhackme.cloud/sites/favicon/images/favicon.ico | md5sum
% Total    % Received % Xferd  Average Speed   Time    Time     Time
Current                                  Dload  Upload  Total   Spent    Left
Speed
0      0     0      0      0      0      0      0  --:--:-- --:--:-- --:--:--
100  1406  100  1406    0      0  20676    0  --:--:-- --:--:-- --:--:--
-- 20676
f276b19aabc4ae8cda4d22625c6735f -
root@ip-10-10-141-115:~#
```

We got the file then we searched on

https://wiki.owasp.org/index.php/OWASP_favicon_database



Answer the questions below

What framework did the favicon belong to?

cgiirc

✓ Correct Answer

💡 Hint

Answer

Task:4

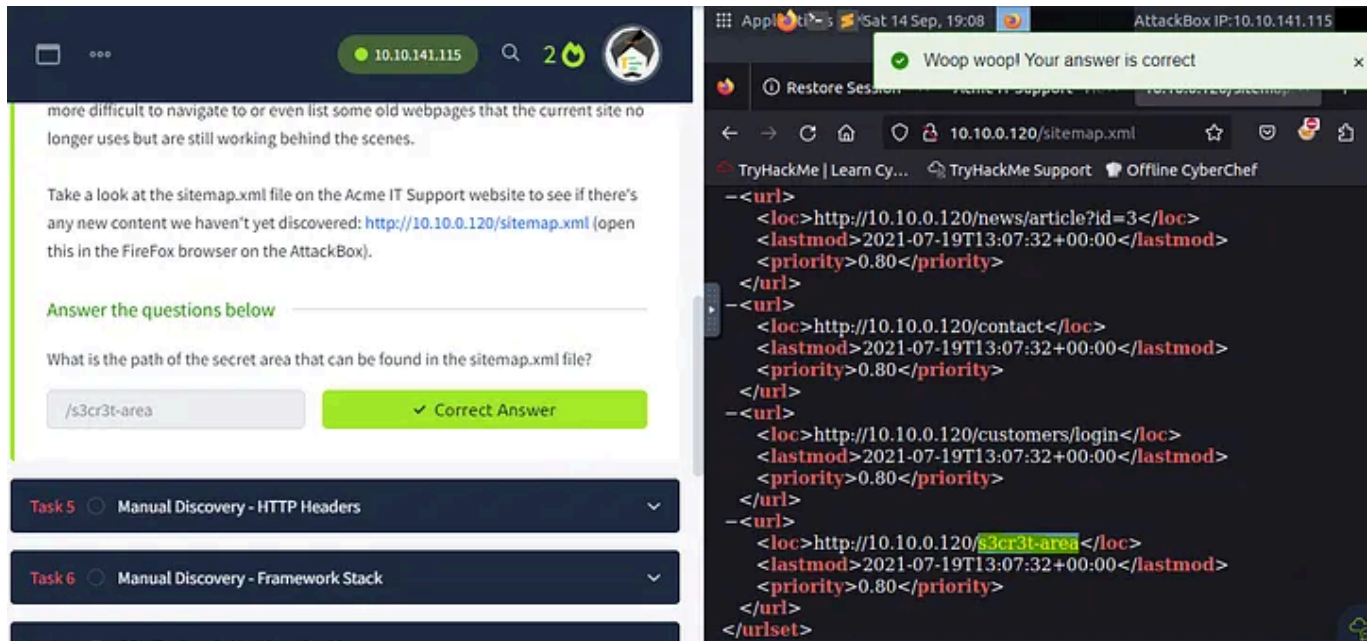
Manual Discovery — Sitemap.xml

Sitemap.xml

Unlike the robots.txt file, which restricts what search engine crawlers can look at, the sitemap.xml file gives a list of every file the website owner wishes to be listed on a search engine.

Take a look at the sitemap.xml file on the Acme IT Support website to see if there's any new content we haven't yet discovered:

<http://10.10.0.120/sitemap.xml>



The image shows two side-by-side screenshots. The left screenshot is from a web application interface with a dark theme. It has a top bar with a search icon, a green status indicator, and a user profile icon. The main content area contains text about navigating to old webpages and a task instruction: "Take a look at the sitemap.xml file on the Acme IT Support website to see if there's any new content we haven't yet discovered: <http://10.10.0.120/sitemap.xml> (open this in the FireFox browser on the AttackBox)." Below this is a question: "What is the path of the secret area that can be found in the sitemap.xml file?" with a text input field containing "/s3cr3t-area" and a green "Correct Answer" button. At the bottom, there are two task cards: "Task 5 Manual Discovery - HTTP Headers" and "Task 6 Manual Discovery - Framework Stack". The right screenshot is a browser window showing the content of the <http://10.10.0.120/sitemap.xml> file. The browser address bar shows the URL. The page content is XML code listing several URLs with their last modified dates and priorities. The URL <http://10.10.0.120/s3cr3t-area> is highlighted in green. A green notification banner at the top of the browser window says "Woop woopl Your answer is correct".

Task:5

Manual Discovery — HTTP Headers

When a web server responds to a request, it sends **HTTP headers** along with the content. These headers can reveal useful details like the web server software and the programming or scripting language used, which may help identify potential vulnerabilities.


```
root@ip-10-10-141-115:~# curl http://10.10.0.120 -v
* Rebuilt URL to: http://10.10.0.120/
* Trying 10.10.0.120...
* TCP_NODELAY set
* Connected to 10.10.0.120 (10.10.0.120) port 80 (#0)
> GET / HTTP/1.1
> Host: 10.10.0.120
> User-Agent: curl/7.58.0
> Accept: */*
>
< HTTP/1.1 200 OK
< Server: nginx/1.18.0 (Ubuntu)
< Date: Sat, 14 Sep 2024 18:10:50 GMT
< Content-Type: text/html; charset=UTF-8
< Transfer-Encoding: chunked
< Connection: keep-alive
< X-FLAG: THM{HEADER_FLAG}
<
<!--
```

we can see the webserver is NGINX version 1.18.0 and runs PHP version 7.4.3. Using this information, we could find vulnerable versions of software being used. Try running the below curl command against the web server, where the -v switch enables verbose mode, which will output the headers

Task:6

Manual Discovery — Framework Stack

Looking at the page source of our Acme IT Support website (<http://10.10.0.120>), you'll see a comment at the end of every page with a page load time and also a link to the framework's website, which is <https://static-labs.tryhackme.cloud/sites/thm-web-framework>. Let's take a look at that website. Viewing the documentation page gives us the path of the framework's administration portal, which gives us a flag if viewed on the Acme IT Support website.


```
view-source:http://10.10.0.120/
e | Learn Cy... TryHackMe Support Offline CyberChef
</a>

<a href="/secret-page">to</a> assist you with your IT problems.

https://static-labs.tryhackme.cloud/sites/thm-web-framework )
```


← → ↻ 🏠 🔒 https://static-labs.tryhackme.clo... ☆ 📧 🍷 📄 ≡

TryHackMe | Learn Cy... TryHackMe Support 🍷 Offline CyberChef >

**Try
Hack
Me**

THM Web Framework

Documentation

[Home](#) • [Change Log](#) • [Documentation](#)

Documentation

The documentation for the framework is pre-installed on your websites administration portal.

Once you've installed the framework navigate to the `/thm-framework-login` path on your website.

You can login with the `username admin` and `password admin` (make sure you change this password)

← → ↻ 🏠 🔒 10.10.0.120/thm-framework ☆ 📧 🍷 📄 ≡

TryHackMe | Learn Cy... TryHackMe Support 🍷 Offline CyberChef >>

THM Web Framework

Login

Login

Username:

Password:

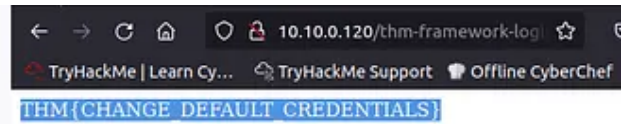
[Login](#)

Answer the questions below

What is the flag from the framework's administration portal?

THM{CHANGE_DEFAULT_CREDI

✓ Correct Answer



Task:7

OSINT — Google Hacking / Dorking

Filter	Example	Description
site	site:tryhackme.com	returns results only from the specified website address
inurl	inurl:admin	returns results that have the specified word in the URL
filetype	filetype:pdf	returns results which are a particular file extension
intitle	intitle:admin	returns results that contain the specified word in the title

More information about google hacking can be found here: https://en.wikipedia.org/wiki/Google_hacking

Answer the questions below

What Google dork operator can be used to only show results from a particular site?

site:

✓ Correct Answer

💡 Hint

Task:8

OSINT — Wappalyzer

Wappalyzer (<https://www.wappalyzer.com/>) is an online tool and browser extension that helps identify what technologies a website uses, such as frameworks, Content Management Systems (CMS), payment processors, and much more, and it can even find version numbers as well.

Answer the questions below

What online tool can be used to identify what technologies a website is running?

Wappalyzer

✓ Correct Answer

Task:9

OSINT — Wavback Machine

Open in app ↗



Medium



Search



Write



Wayback Machine

The Wayback Machine (<https://archive.org/web/>) is a historical archive of websites that dates back to the late 90s. You can search a domain name, and it will show you all the times the service scraped the web page and saved the contents. This service can help uncover old pages that may still be active on the current website.

Answer the questions below

What is the website address for the Wayback Machine?

<https://archive.org/web/>

✓ Correct Answer

Task:10

OSINT — GitHub

What is Git?

version control system

✓ Correct Answer

Task:11

OSINT — S3 Buckets

tryhackme-assets.s3.amazonaws.com

What URL format do Amazon S3 buckets end in?

.s3.amazonaws.com

✓ Correct Answer

💡 Hint

Task:12

Automated Discovery

What is Automated Discovery?

Automated discovery is the process of using tools to discover content rather than doing it manually. This process is automated as it usually contains hundreds, thousands, or even millions of requests to a web server. These requests check whether a file or directory exists on a website, giving us access to resources we didn't previously know existed