

Google Dorking

TryHackMe Dashboard Learn Practice Compete

Learn > Google Dorking

Google Dorking

Explaining how Search Engines work and leveraging them into finding hidden content!

45 min 1,36,870

Help Save Room 5178 Recommend

Room progress (0%)

To access material, start machines and answer questions you need to join this room!

Join Room

Google Dorking | DarkStar • Jan 17, 2021

Source: YouTube

Room progress (25%)

A logical presumption would be that website 1 -> 3 would be displayed...But that's not how real-world domains work and/or are named.

So, who (or what) decides the hierarchy? Well...

Answer the questions below

Name the key term of what a "Crawler" is used to do. This is known as a collection of resources and their locations

Index ✓ Correct Answer

What is the name of the technique that "Search Engines" use to retrieve this information about websites?

Crawling ✓ Correct Answer

What is an example of the type of contents that could be gathered from a website?

Keywords ✓ Correct Answer

Task 3 Enter: Search Engine Optimisation

Task 4 Beepboop - Robots.txt

tryhackme.com/room/googleclorking

Room progress (25%)

Sign in

According to this tool, TryHackMe has an SEO rating of **85/100** (as of 14/11/2020). That's not too bad and it'll show the ju

Woop woopl! Your answer is correct

But...Who or What Regulates these "Crawlers"?

Aside from the search engines who provide these "Crawlers", website/web-server owners themselves ultimately stipulate what content "Crawlers" can scrape. Search engines will want to retrieve **everything** from a website - but there are a few cases where we wouldn't want **all** of the contents of our website to be indexed! Can you think of any...? How about a secret administrator login page? We don't want **everyone** to be able to find that directory - especially through a google search.

Introducing Robots.txt...

Answer the questions below

Use the same [SEO checkup tool](https://tryhackme.com) and other online alternatives to see how their results compare for <https://tryhackme.com> and <http://googleclorking.cmndtic.co.uk>

No answer needed

Correct Answer

Task 4 Beepboop - Robots.txt

Task 5 Sitemaps

tryhackme.com/room/googleclorking

Room progress (56%)

Sign in

Where would "robots.txt" be located on the domain "ablog.com"

ablog.com/robots.txt

Correct Answer

Hint

If a website was to have a sitemap, where would that be located?

/sitemap.xml

Correct Answer

How would we only allow "Bingbot" to index the website?

User-agent: Bingbot

Correct Answer

How would we prevent a "Crawler" from indexing the directory "/dont-index-me/"?

Disallow: /dont-index-me/

Correct Answer

What is the extension of a Unix/Linux system configuration file that we might want to hide from "Crawlers"?

.conf

Correct Answer

Hint

Task 5 Sitemaps

tryhackme.com/room/googledorking Room progress (81%)

The easier a website is to "Crawl", the more optimised it is for the "Search Engine"

✓ Woop woop! Your answer is correct

Answer the questions below

What is the typical file structure of a "Sitemap"?

XML ✓ Correct Answer

What real life example can "Sitemaps" be compared to?

Map ✓ Correct Answer

Name the keyword for the path taken for content on a website

Route ✓ Correct Answer

Task 6 What is Google Dorking?

How likely are you to recommend this room to others?

tryhackme.com/room/googledorking Room progress (93%)

✓ Woop woop! Your answer is correct

Answer the questions below

What would be the format used to query the site bbc.co.uk about flood defences

site: bbc.co.uk flood defences ✓ Correct Answer Hint

What term would you use to search by file type?

filetype: ✓ Correct Answer

What term can we use to look for login pages?

intitle: login ✓ Correct Answer Hint

How likely are you to recommend this room to others?

