**Google Dorks**

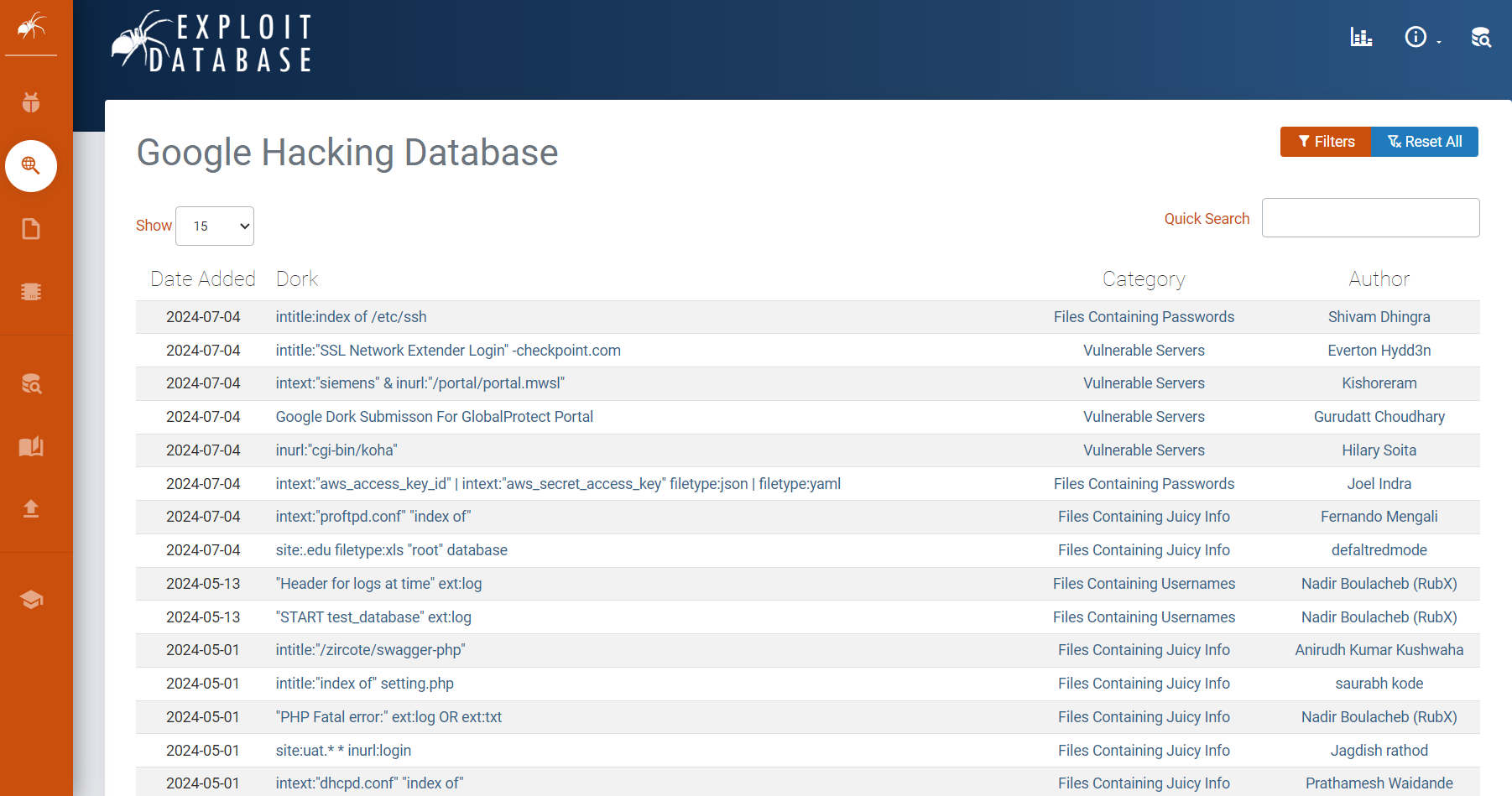
A Google Dork, also referred to as Google Dorking or Google hacking, is a valuable tool for security researchers. While the average person uses Google to find text, images, videos, and news, in the infosec world, Google can be a powerful hacking tool.

You can't hack websites directly using Google, but its extensive web-crawling capabilities can index almost anything on your website, including sensitive information.

We can use special arguments in a normal google query to find specific information. Dork comes in a format **operator:keyword.**

Google's web crawlers will index all publicly accessible content on your website by default. To prevent specific resources from being indexed, you need to use a robots.txt file to explicitly block them. Without such restrictions, anyone using appropriate search queries can potentially find and access the indexed information.

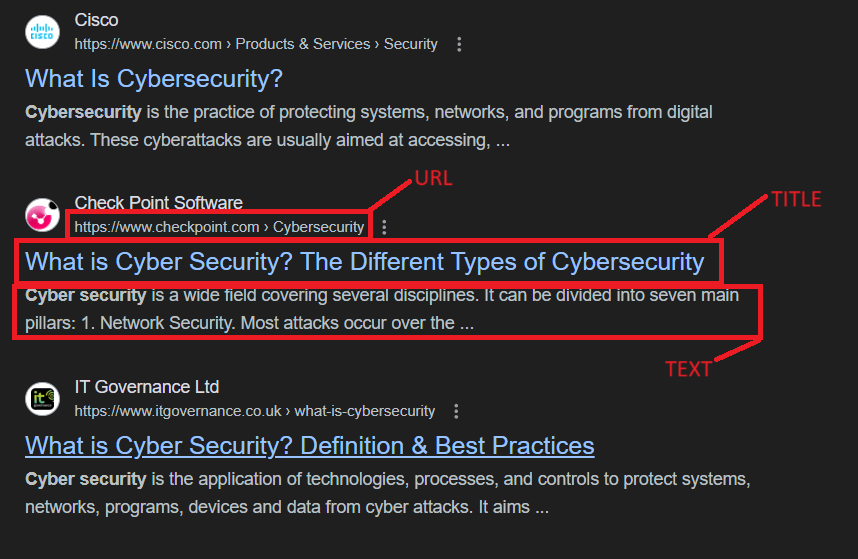
The **Google Hacking Database (GHDB)** lists all Google dorking commands.



It's important to understand that although this information is publicly accessible and legal to use, it can be misused by malicious individuals to damage your online presence.Google tracks your activity when performing these queries, so it's advised to use them responsibly, either for your own research or to defend your website against vulnerabilities.

Google Dorks are search queries that can help you find specific information on the web by exploiting the advanced search operators in Google Search.

1. **URL** is the web address of a website.
2. **text**: refers to the content or body of a webpage
3. **Title** is the webpage’s name or heading which is often displayed in the browser's title bar and search engine results



**3 most important use cases are**

1. **Retrieving files from domains**

**“**Cyber security filetype:pdf”

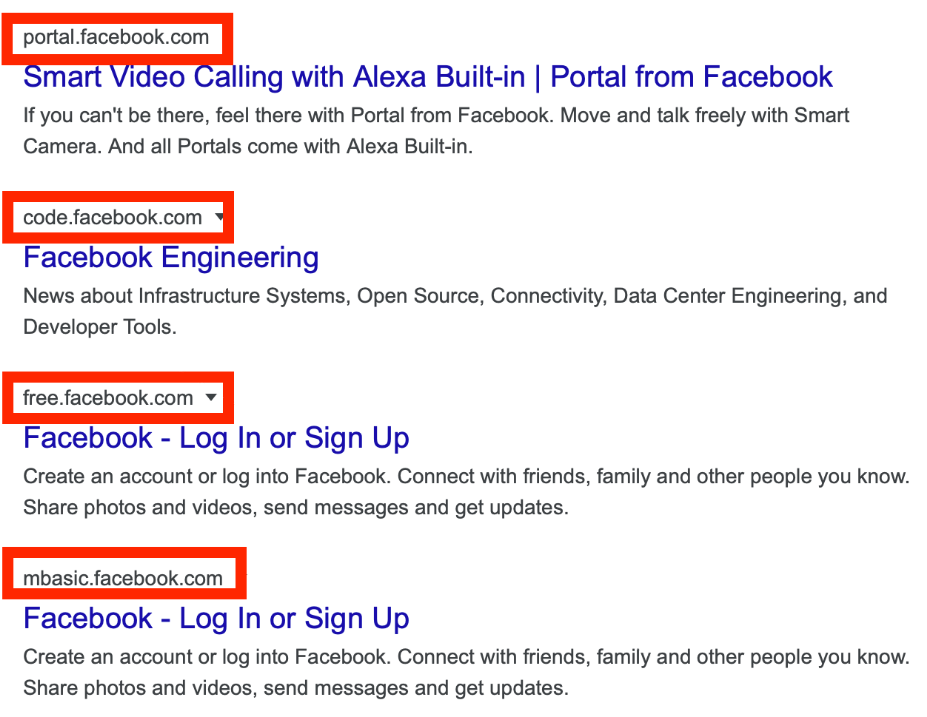


In the above screenshot we can see that the search has in fact brought back any PDF files that contain the strings “cyber” and “security”. We can use this to see what files a company is hosting online, and see if any are confidential and should not be publicly accessible. It is also possible to retrieve information about internal systems and users by looking at the metadata of files to see when they were created and who by. Documents can also be used to create custom wordlists for password attacks against specific organizations.

1. **Subdomian enumeration**

“site:facebook.com -site:www.facebook.com”

It means to look for sites that include .facebook.com but exclude www.facebook.com.

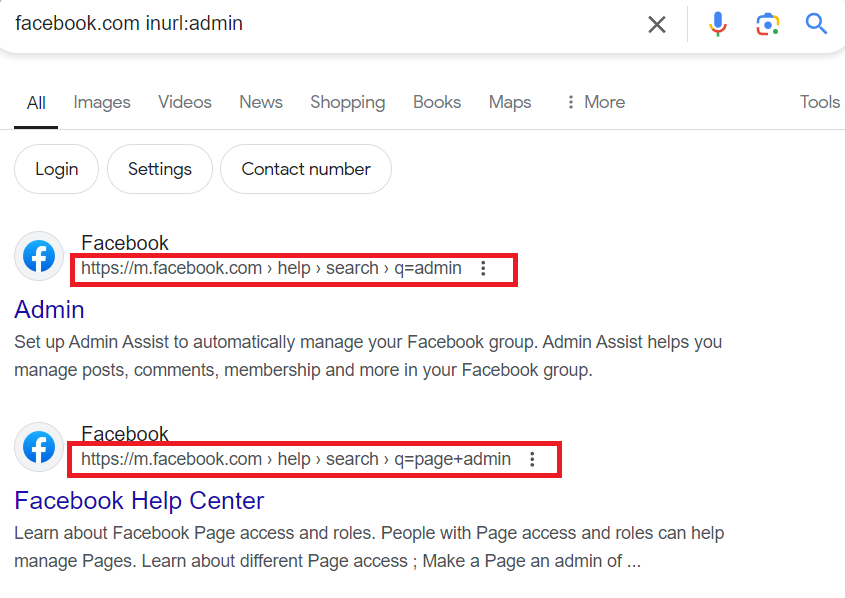


Here we can see the list begins with two subdomains, portal.facebook.com, and code.facebook.com. We have successfully enumerated subdomains using Google Dorks. This is a great method for identifying uncommon web pages that may feature a login portal or valuable information such as development environments, files, and more. Have a go yourself with any domain you want, and see what interesting subdomains they have!

1. **Keyword Searching**

“facebook.com inurl:admin”

Look for the admin keyword in the url of facebook.



Using the dork inurl:value we can look for specific keywords in a much more refined way than normal google searches. using the query inurl:admin we can see a number of what appears to be admin login portals. This would be great if we were working as an attacker(if in scope, we can brute force or bypass the login portal to access administrator dashboards) or a defender(we can work to secure these portals so they are not compromised.

Apart from important use cases let's look into the list of other search filters

| **Filter** | **Description** | **Example** |
| --- | --- | --- |
| allintext | Searches for occurrences of all the keywords given. | allintext:"keyword" |
| intext | Searches for the occurrences of keywords all at once or one at a time. | intext:"keyword" |
| allinurl | Searches for a URL matching all the keywords in the query. | allinurl:"keyword" |
| intitle | Searches for occurrences of keywords in title all or one. | intitle:"keyword" |
| link | Searches for external links to pages. | link:"keyword" |
| related | List web pages that are “similar” to a specific web page. | related:www.google.com |
| cache | Shows the version of the web page that Google has in its cache. | cache:www.google.com |

**Defense against google dorks**

We know that Google Dorks can be incredibly powerful – they can expose admin login portals, usernames and passwords, IP cameras and webcams, and much more. But how exactly do we protect against them?

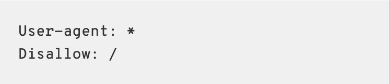
1. **Geofencing and IP whitelisting**

IP-based controls can restrict access to web content, blocking unauthorized users or Google's crawlers. Geofencing blocks entire IP ranges from specific countries, allowing access only from a desired location. However, this method has a security flaw: users from other countries can bypass it using a VPN set to exit from the allowed country. For example, if a UK business restricts access to UK IP addresses, someone in America could use a VPN with a UK exit point to gain access. Netflix uses IP geofencing to manage different regional content versions, but users initially bypassed this with VPNs. Netflix then started blocking VPN and proxy connections, prompting users to disable them to access content.

IP whitelisting, on the other hand, allows only specified IP addresses to access resources and blocks all others. This is useful for development environments or restricted sites on the internet, ensuring that only IPs from a specific organization can view the site.

2. **Crawler restrictions**

If you wish to keep certain content on your website from appearing in Google's search results, you can implement a robots.txt file.By configuring the robots.txt file to disallow all crawlers from indexing any part of your site, you can ensure your content won't appear in Google's search results.



This configuration disallows any user-agent (represented by \*) from crawling any directory on the website, such as /images/ or /training/.

3.  **Requesting content removal**

If you run google dorks against your company and find results that could be used by a malicious actor (such as a login portal or sensitive files) you can make a request to Google asking them to temporarily remove the content from their search engine (90 days) or permanently remove it. You must provide sufficient evidence that you own the site in order to have anything removed.

In conclusion, while Google Dorks offer powerful capabilities for both attackers and defenders in uncovering vulnerabilities and securing online assets, their potential for misuse underscores the importance of proactive defense measures. Implementing strategies like geofencing, IP whitelisting, and crawler restrictions can significantly mitigate the risks associated with unintentional exposure of sensitive information.