Lab - Advanced Searches

Objectives

Use Google Advanced Search to perform passive reconnaissance.

- Part 1: Google Advanced Searches (Dorking)
- Part 2: The Google Hacking Database
- Part 3: The Wayback Machine

Background / Scenario

The first step a hacker takes is to learn as much information about a target as possible. The more the attackers know about the target, the better they can hack it with other hacking techniques. Using Advanced Google searches and parsing through archived internet sites are two popular methods of passive reconnaissance. They help inform ethical hackers about a client's vulnerabilities and pave the way for exploitation activities if they are part of the scope of the test.

Through an advanced Google search, the hacker is hoping to find information that has been made public by accident. For example, someone may have accidentally exposed passwords, left a webcam open to the internet, or revealed other useful information. The hacker will search using specific key words and Google search operators to try and find what they are looking for. This is called **Google dorking**. It involves using specific Google search queries to uncover information that was not meant to be publicly available.

The **Wayback Machine web archive** is another useful tool for uncovering potential vulnerabilities. Valuable personal and corporate information can sometimes be gleaned from archived web pages. Using the Wayback machine, a hacker can browse through the history of a website and visit snapshots of the site at various times in the past. This allows the hacker to uncover information no longer available on the live internet that may be useful for further attacks.

Unauthorized access to data, computers, and network systems is a crime in many jurisdictions and often is accompanied by severe consequences, regardless of the perpetrator's motivations. It is the learner's responsibility, as the user of this material, to be cognizant of, and compliant with, computer use and privacy laws.

Required Resources

- Computer or mobile device with internet access

Instructions

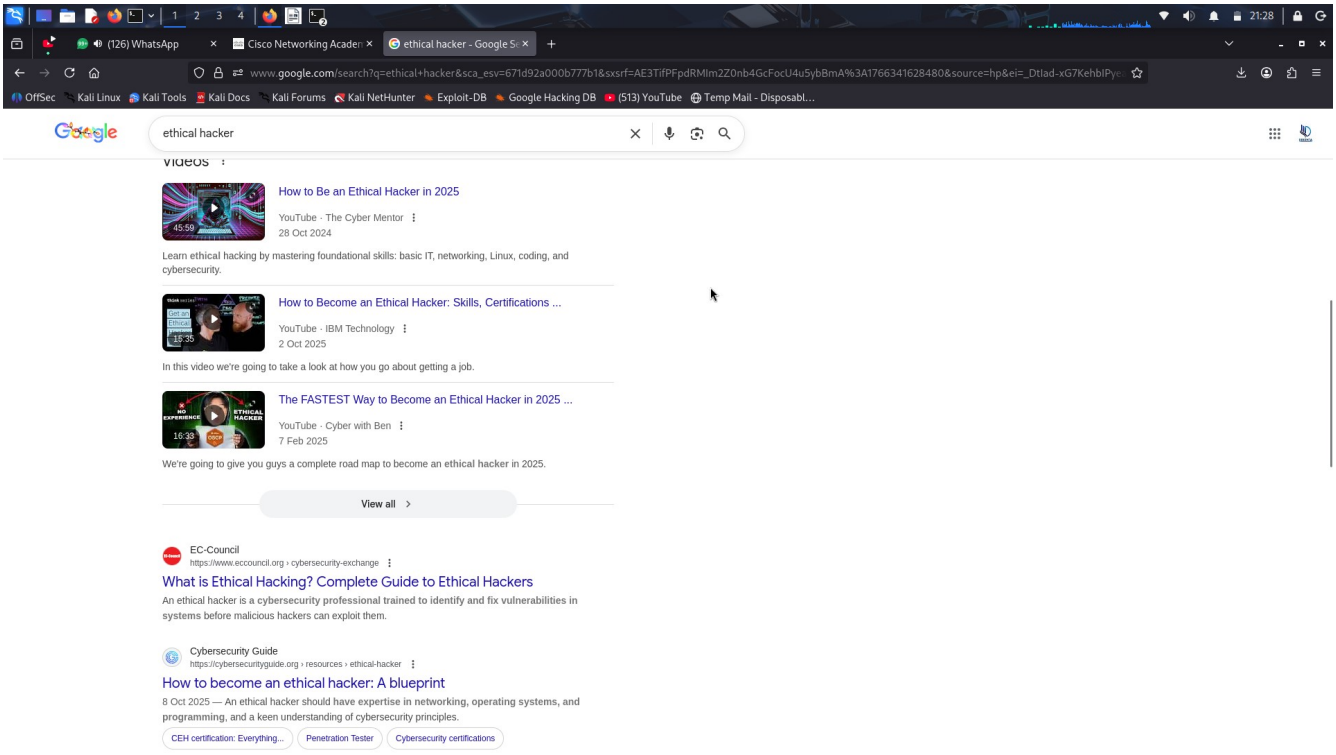
Part 1: Google Advanced Searches (Dorking)

For most people, Google is a tool for searching for text, images, videos, and news on the internet by using simple string queries. However, for some, Google is a powerful and useful hacking tool and can be used for performing passive reconnaissance by using advanced search operators. The practice of using advanced Google search operators to find information and vulnerable servers is called **Google dorking or Google hacking.** Google dorking is used by hackers to try to find information that was never intended to be revealed publicly. It is a useful technique for conducting passive reconnaissance in penetration tests.

Note: When performing advanced queries, you may be prompted by Google to prove you are not a robot. If this occurs, as it probably will after several searches, simply complete the captcha and continue.

Step 1: Explore Google dorking.

a. Navigate to www.google.com to open the Google search engine.
b. Type the string query **ethical hacker** in the search window. Scroll through the results.



c. Note the variety of results returned. This is how we typically use Google to perform searches. String queries like this return a lot of results. However, about 90% of the results are not specific to what we are after. To restrict results to only what is desired, such as pages from a single site, specific keywords, or specific file types, Google Advanced Search operators can be used.

d. There are many Google advanced search operators. Lists of them are available on the internet on sites such as SpyFu. Search the internet for “advanced search operators” to see other source of information, some of which have useful examples.

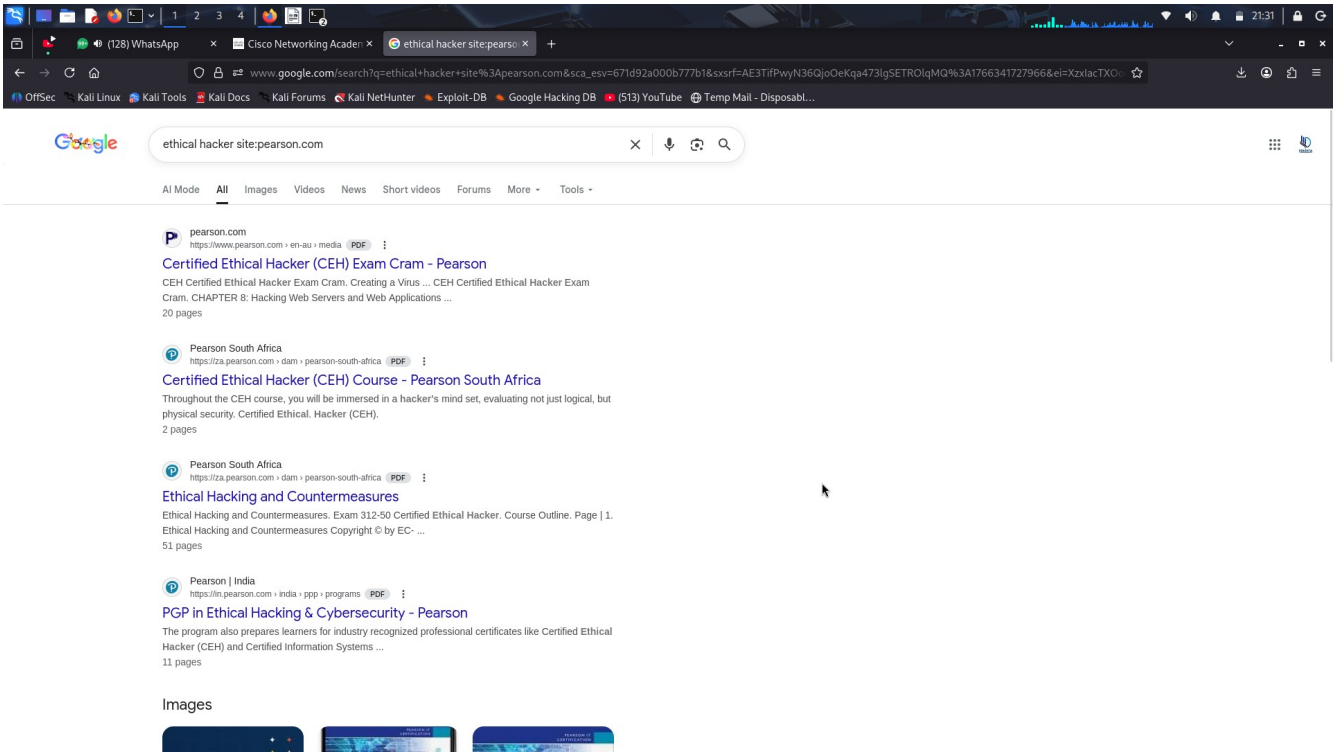
The table below shows the advanced search operators that are used in this lab.

Operator	Description
allintext:	Restricts results to pages with all query words in the page text.
filetype:	Restricts results to pages of the specified file type (.pdf, .ppt, .doc, etc.)
intitle:	Restricts results to pages with a certain word (or words) in the title.
inurl:	Restricts results to pages with a certain word (or words) in the URL.
site:	Restricts results to pages from the specified domain.

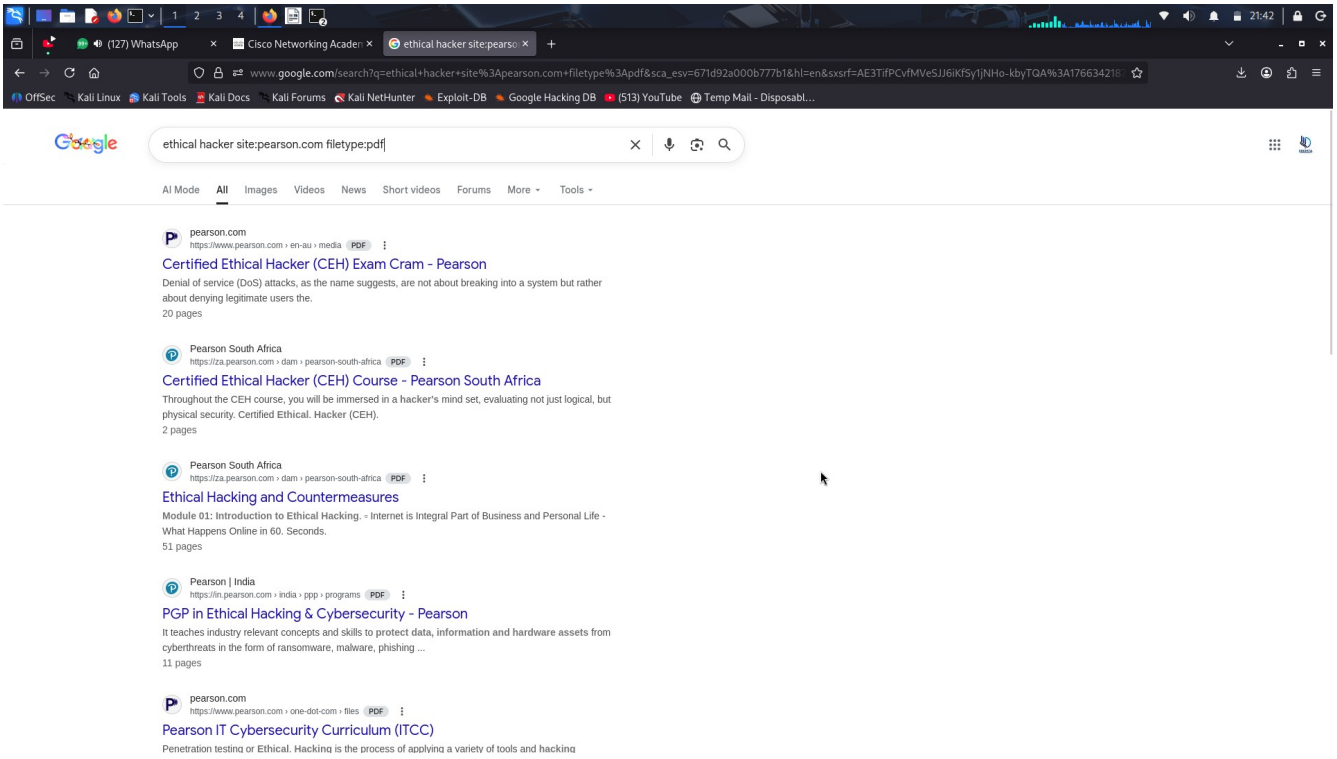
the SCORE function.

Try each of the operators in a Google search. When using advanced search, don’t put spaces between the operator and the domain or keywords.

c. Type **ethical hacker site:pearson.com** in the search window. The syntax is **search term operator:domain**. Scroll through the results.

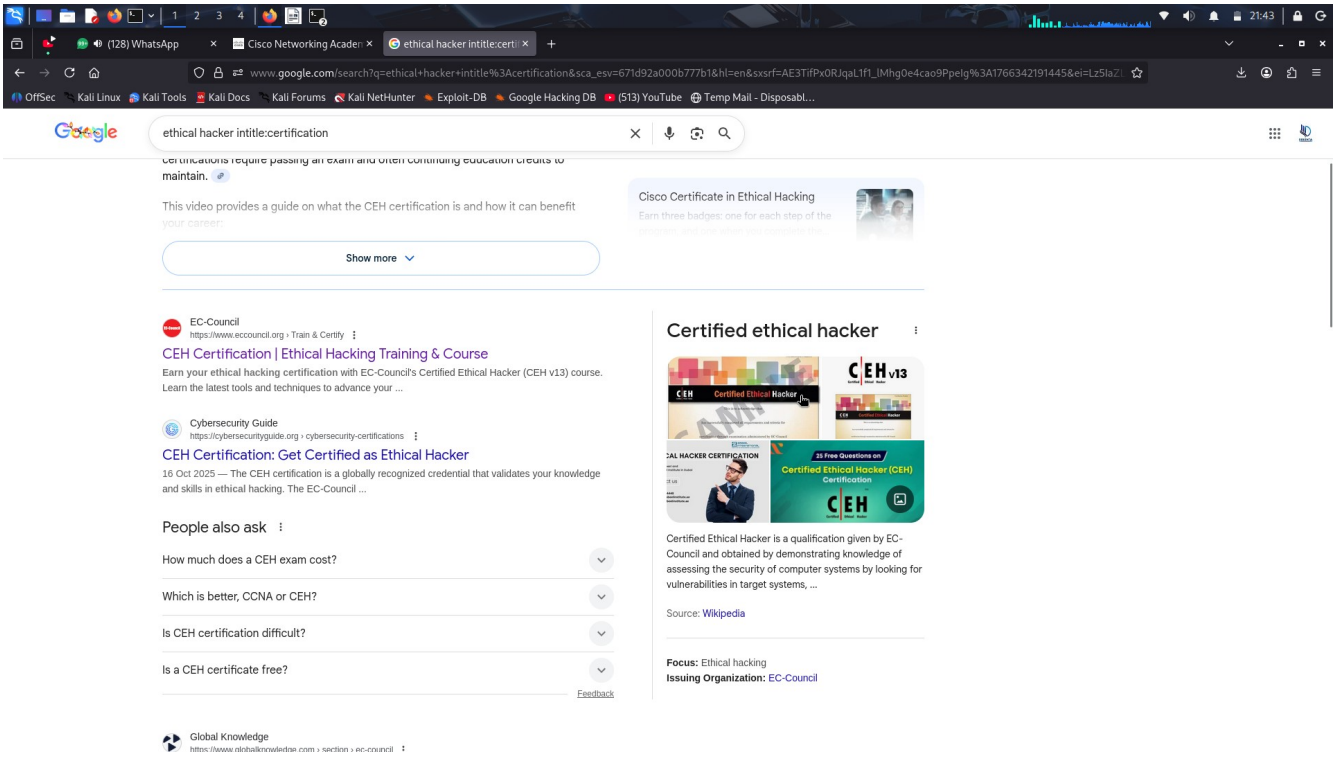


Type **ethical hacker site:pearson.com filetype:pdf** in the search window. Scroll through the results.



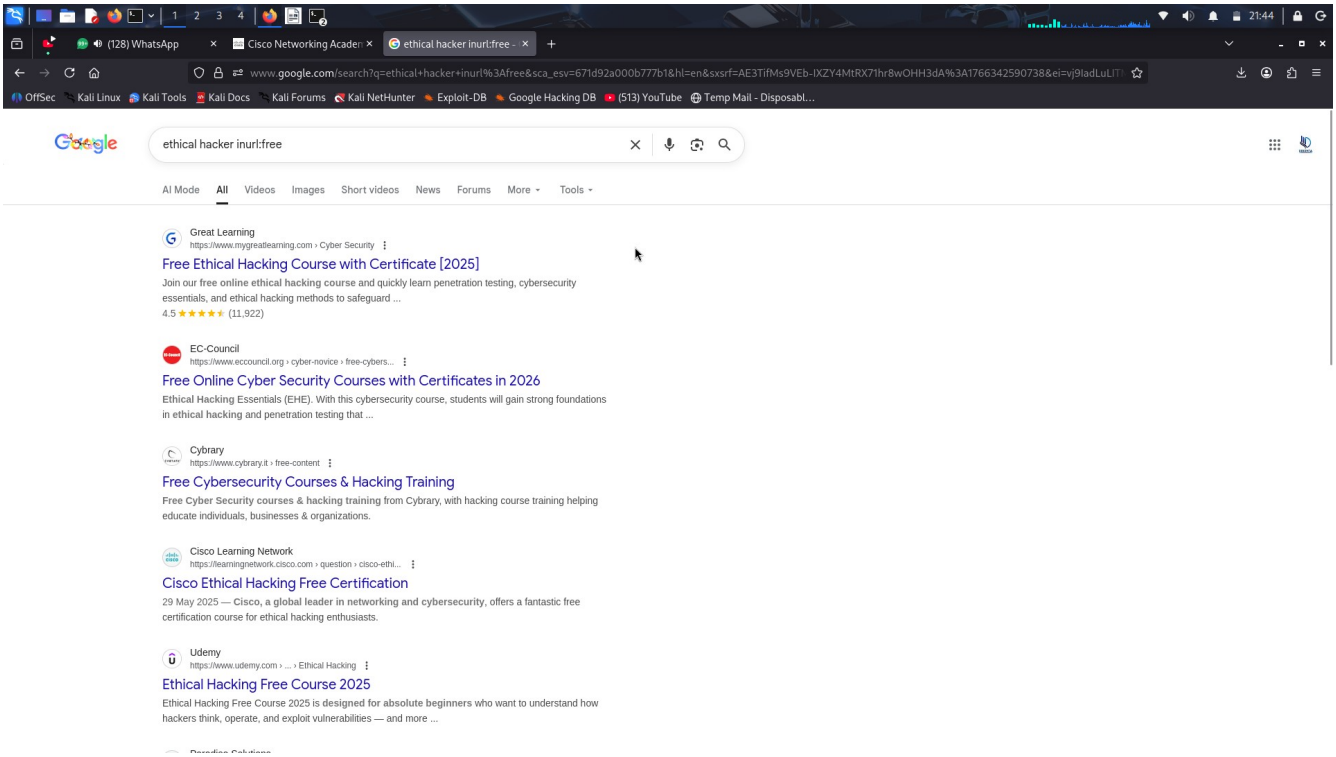
Type **ethical hacker intitle:certification** in the search window. Scroll through the results.

All the results should be related to ethical hacking and include the keyword **certification** in the page title.



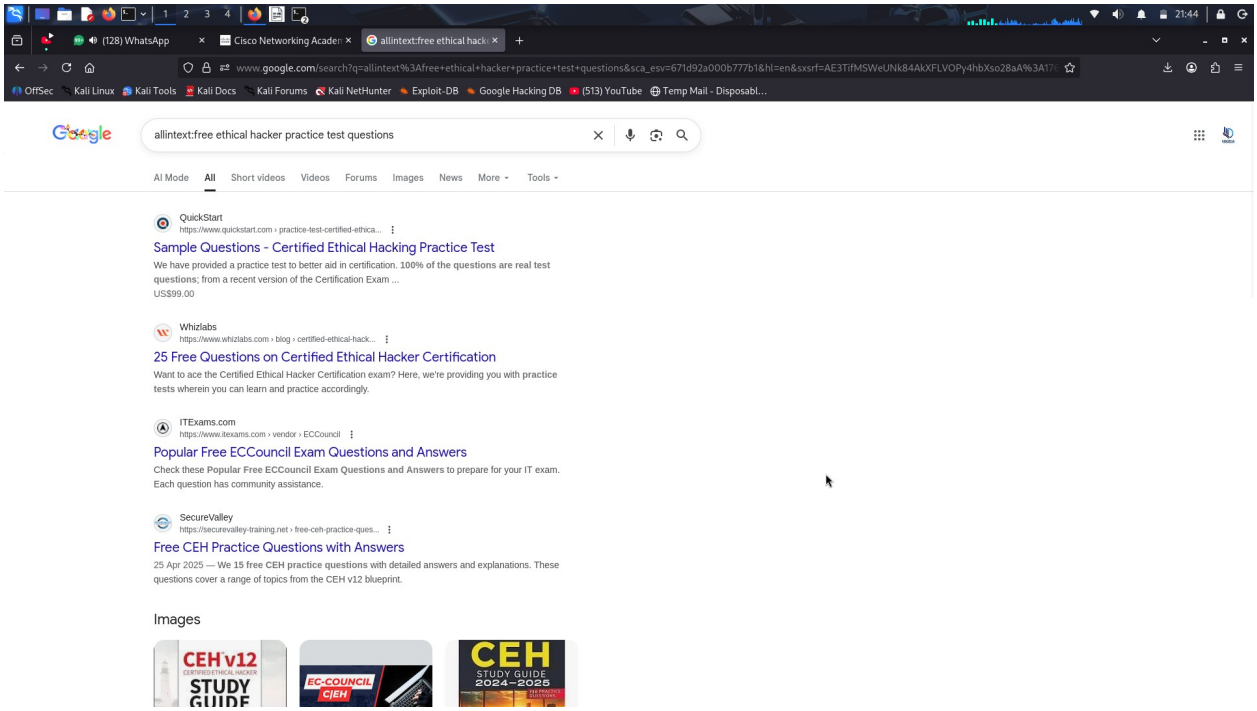
f. Type **ethical hacker inurl:free** in the search window. Scroll through the results.

All the results should be related to ethical hacking and should have the keyword **free** in the URL.



g. Type **allintext:free ethical hacker practice test questions** in the search window. This performs virtually the same function as a normal Google search, but it only returns results with every keyword in the page text. It won't return results with the keywords in only the title. Try putting quotes around your search text.

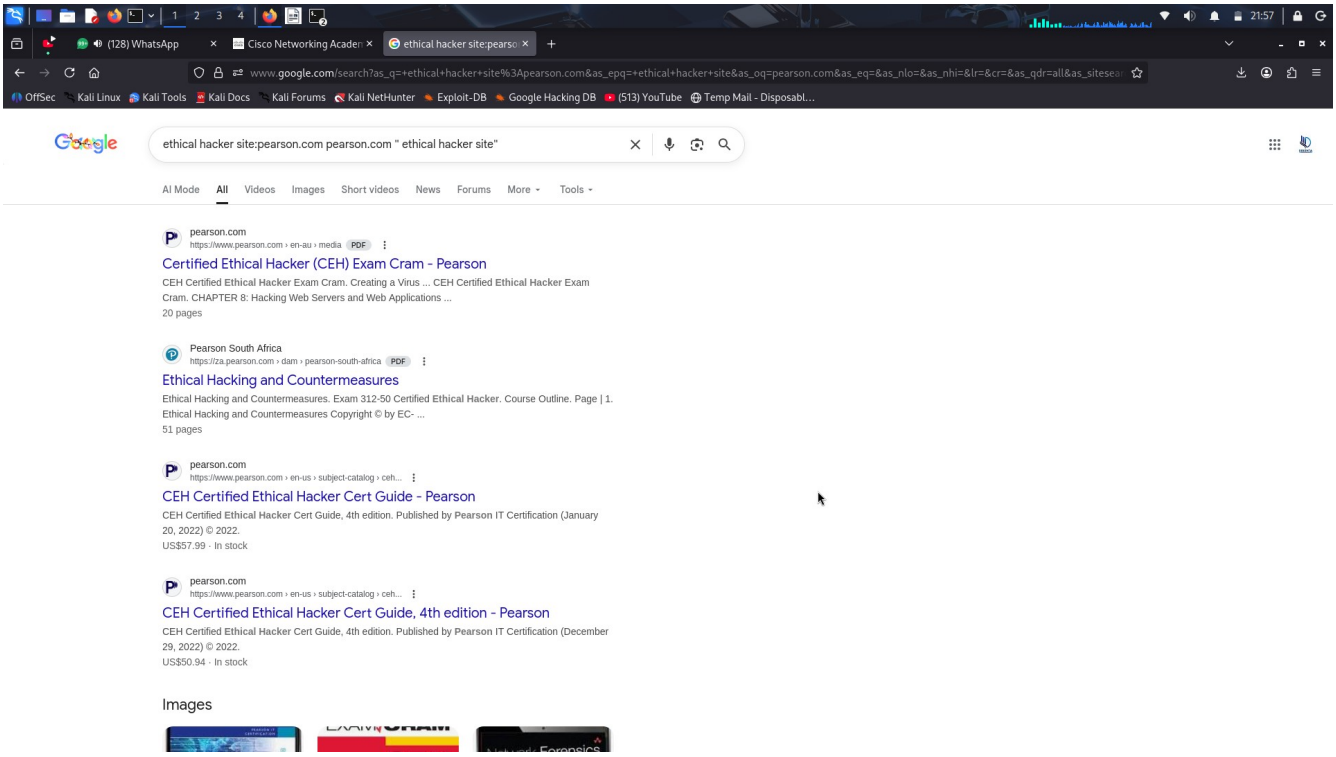
The results should include all the keywords in the page text.



Step 2: Conduct searches using the Google Advanced Search form.

The Google Advanced Search form offers the same result filtering functionality as the common text operators.

a. Type **advanced search** in the Google search window. This will return a link to the advanced search form.
b. Use the advanced search form to perform the same searches that were conducted in the previous step.



Step 3: Conduct passive reconnaissance with advanced search operators.

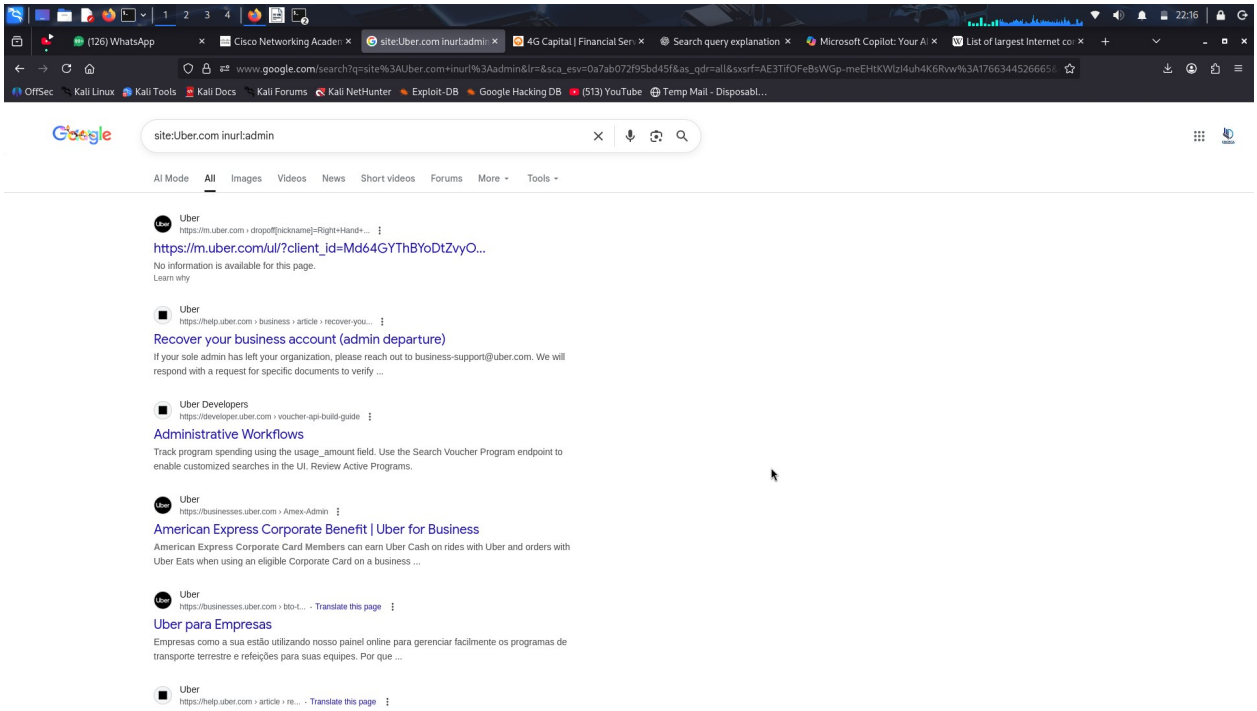
Advanced search operators are useful for narrowing down search results as you have seen. This makes them useful for performing passive reconnaissance as well. Hackers will use advanced search operators to find vulnerabilities and information about potential targets. While the results of the searches may seem harmless on their own, when pieced together, they can provide valuable intelligence to a hacker. The hacker hopes to find sites or files that the target company did not intend to make public, or to find information that can be used for future attacks, such as social engineering attacks.

When performing these searches, use a target company of your choice. Passive reconnaissance is legal but stop there because using any information you uncover for active reconnaissance is not. If you do find vulnerabilities, consider informing the company so that they can correct the issue.

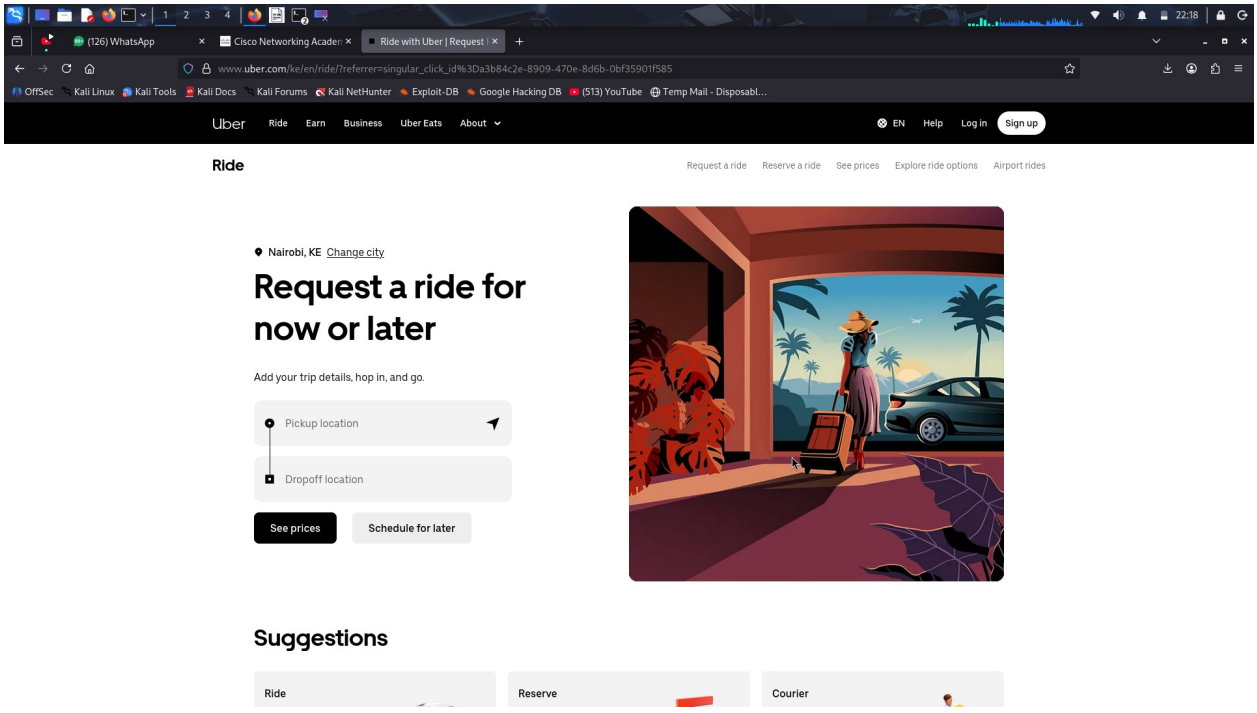
a. Search the target company site using the **inurl:** operator.

In the search window type the command **site:*examplecompany.com* inurl:admin** replacing *examplecompany.com* with a company of your choice.

This will return pages that have the keyword **admin** somewhere inside the URL.



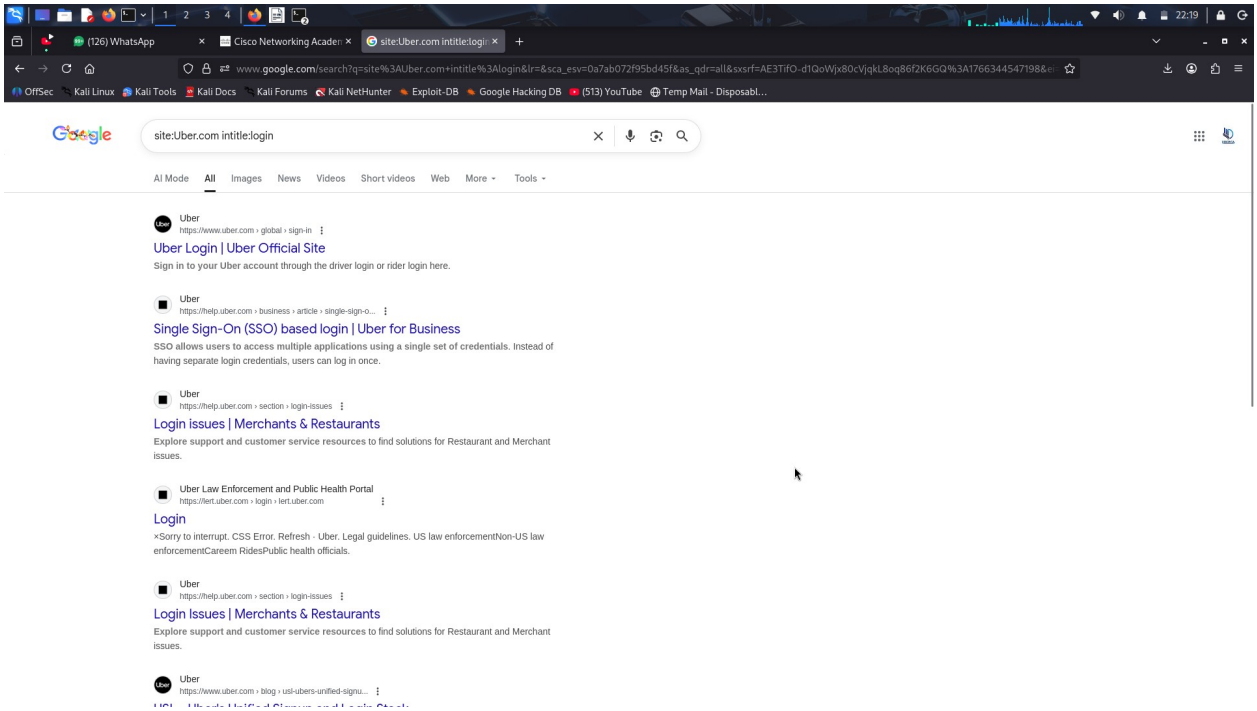
Review the returned pages and click a few to see if there is any interesting information.



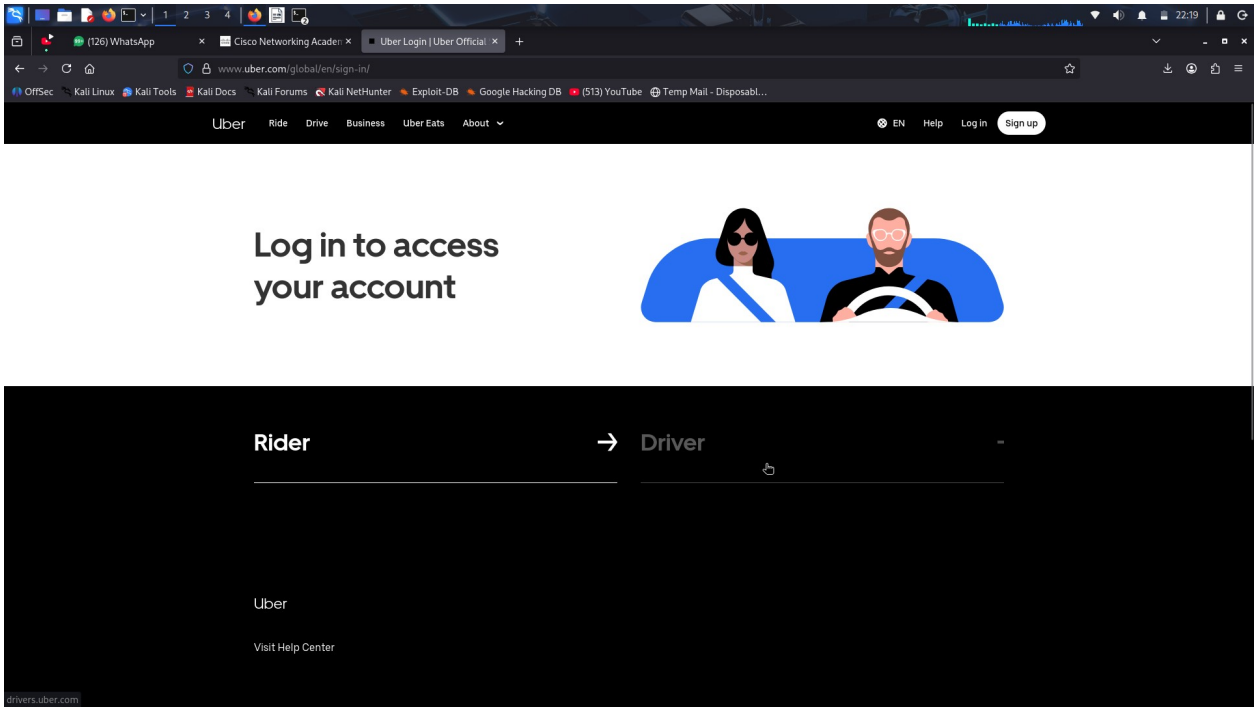
b. Do another search, this time using the **intitle:** operator.

In the search window type the command **site:*examplecompany.com* intitle:login**.

This will return pages that have the keyword **login** in the title.



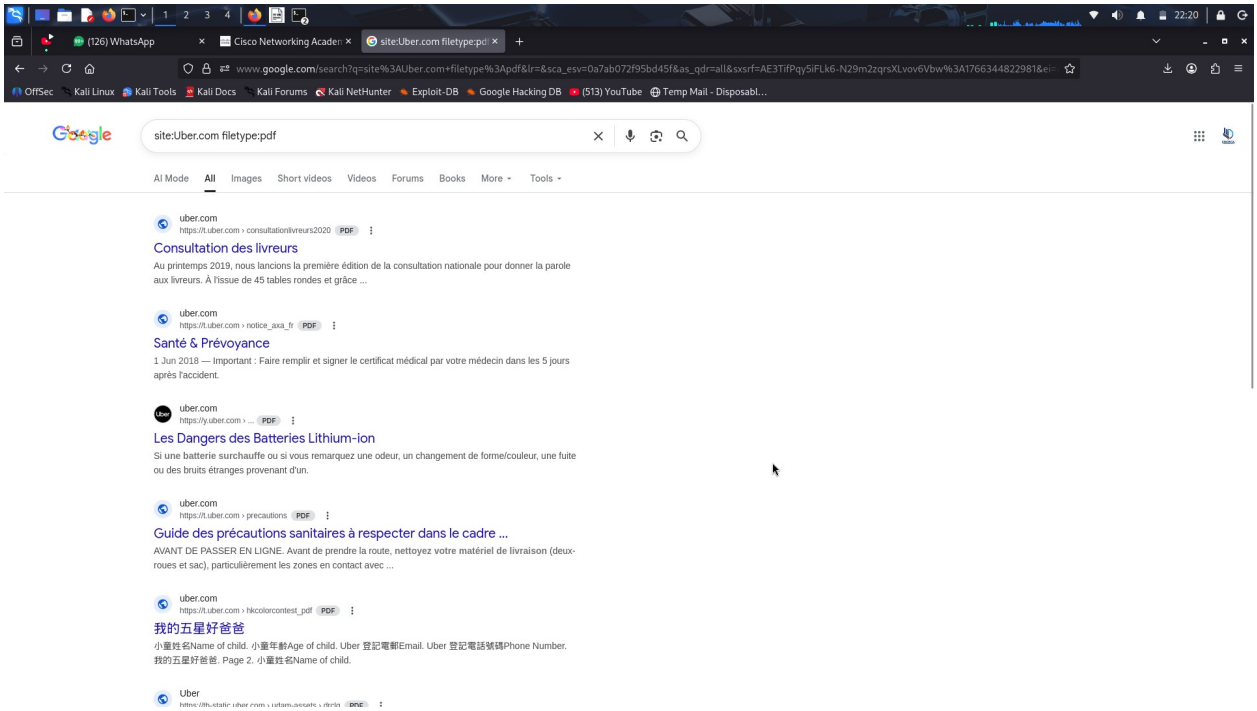
Again, review the results and click a few to see if there is any interesting information.



c. Next, try using the **filetype:** operator.

In the search window, type the command **site:*examplecompany.com* filetype:pdf**.

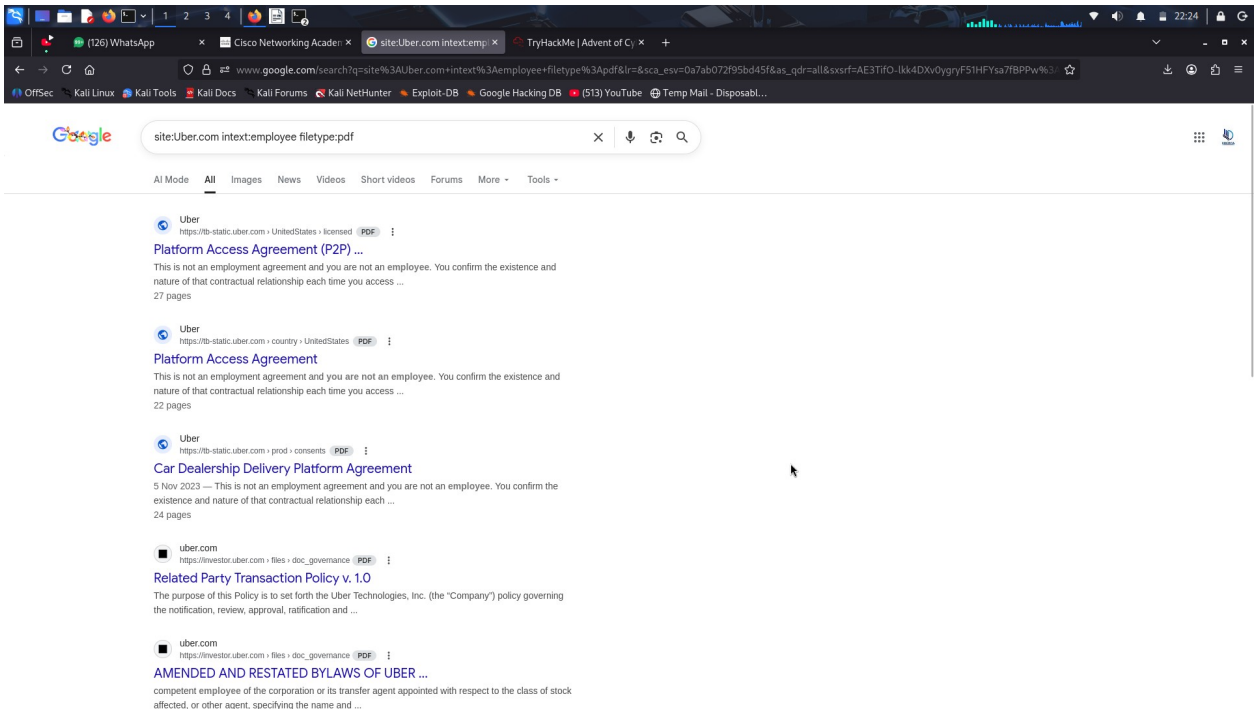
This will return PDF files.



Review some of the files to see if there is any interesting information that is not intended for public access or is useful for social engineering attacks.

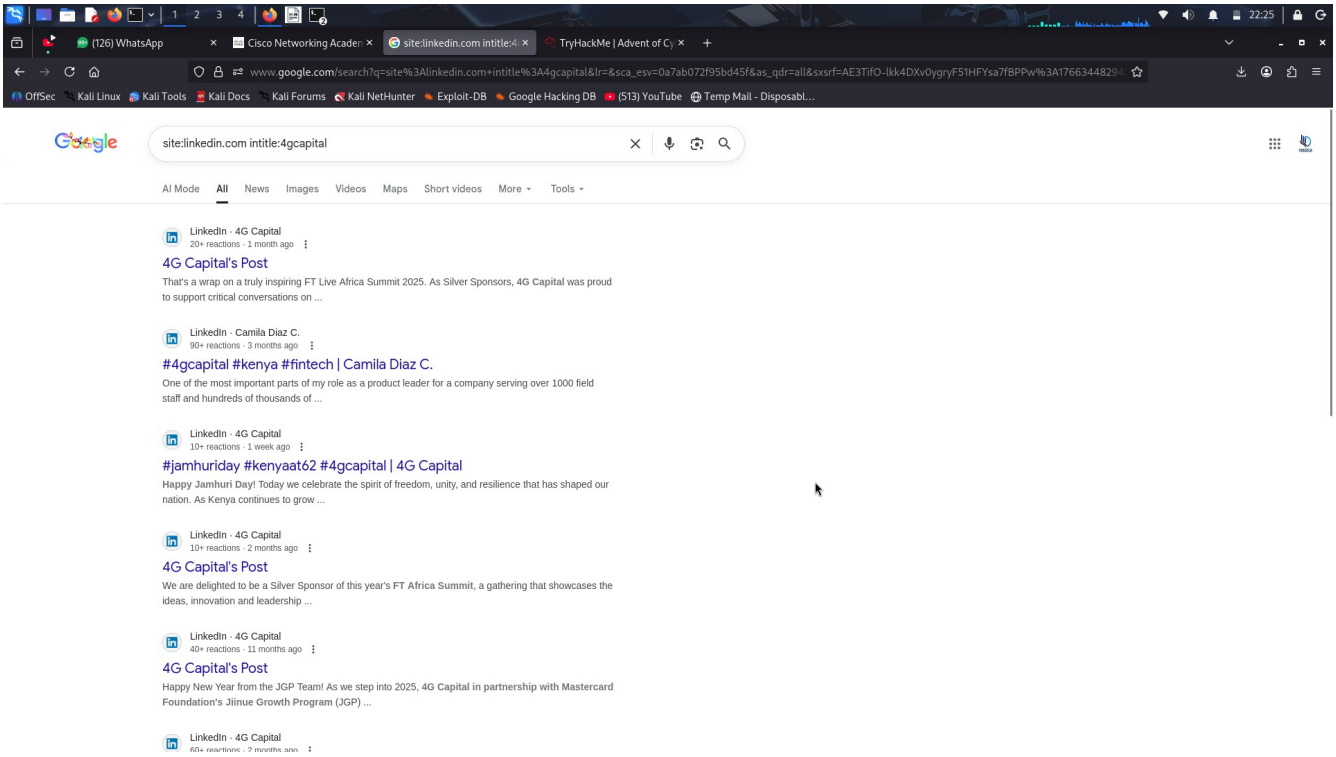
d. Try a search with multiple operators. Use the **intext:** and **filetype:** operators. In the search window, type the command **site:*examplecompany.com* intext:employee filetype:pdf**

This will return PDF pages containing the text **employee**.

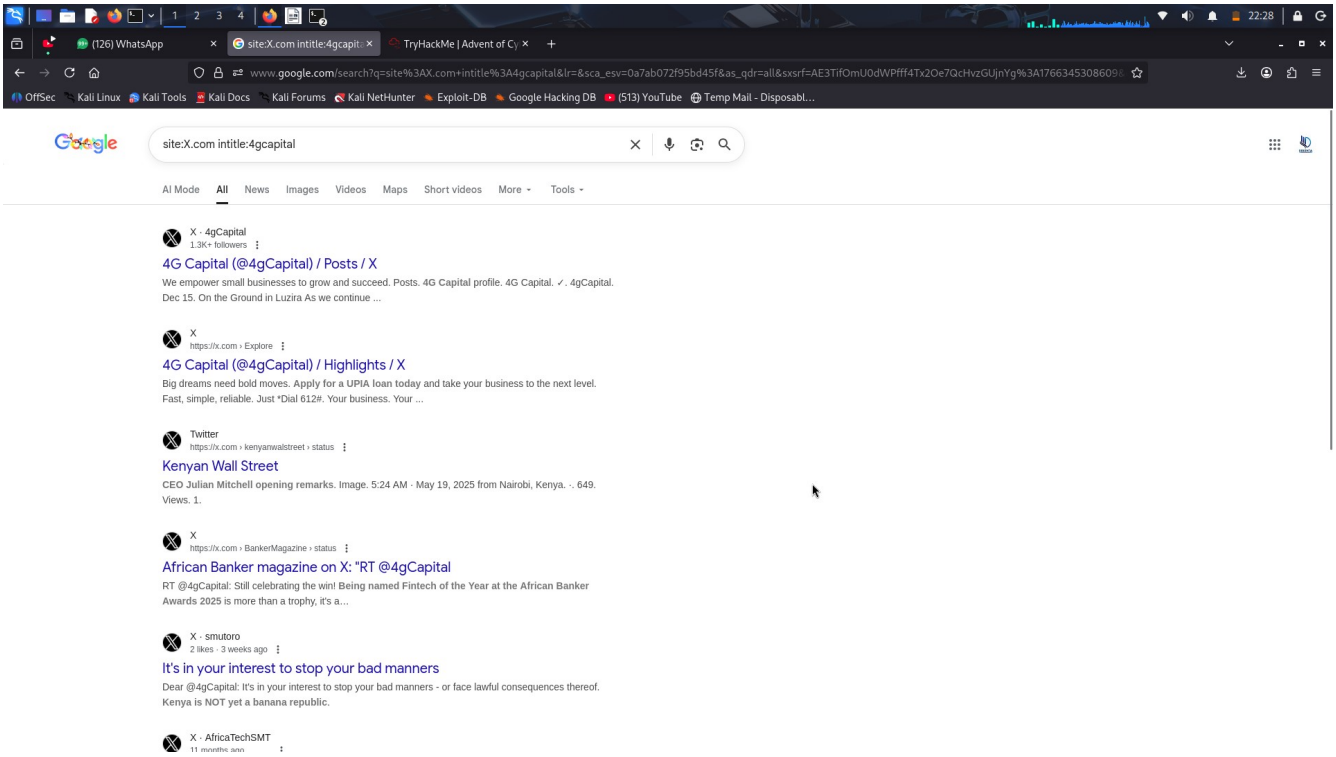


Experiment with **site:*examplecompany.com* intext:<keyword> filetype:<file type>** using different key words and different filetypes.

- LinkedIn can offer valuable information about a company and employees. In the search window, type the command **site:linkedin.com intitle:*example company***. Experiment by searching for the company name with and without the .com at the end.



Experiment with **site:<*social media site*> intitle:*example company*** and search other social media sites.



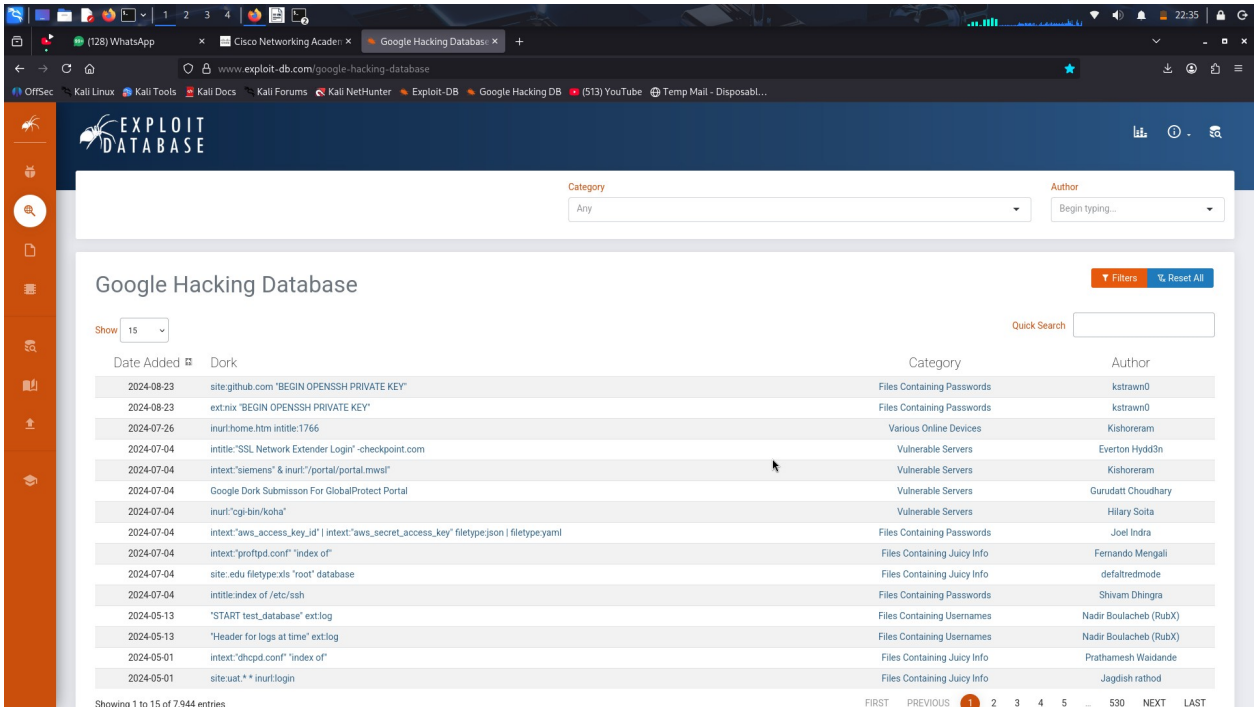
Part 2: The Google Hacking Database

The **Google Hacking Database (GHDB)** is an index of user-created dorks that are designed to uncover interesting, and potentially sensitive, information that was unintentionally made publicly available on the internet.

Step 1: Explore the Google Hacking Database main page.

a. Do a Google search for **GHDB**. The first returned page should be The Google Hacking Database.
b. On the GHDB main page, click the **Filters** button in the top right of the window.

This allows you to filter the database results by Category or Author. There is also a **Quick Search**.

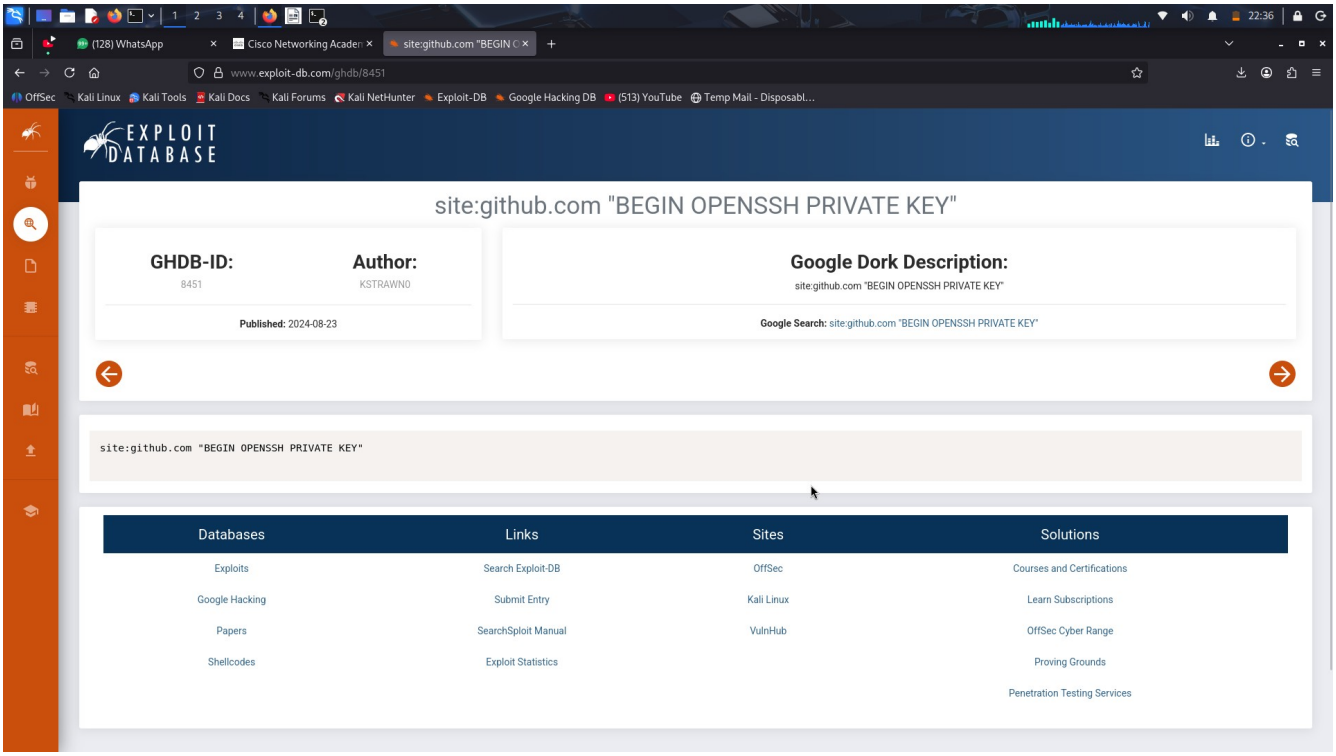


Step 2: Use Quick Search to find specific dorks.

a. Select each of the filter categories and observe some of the dorks available in that category. Select a few interesting looking dorks in the results and note the descriptions of each.

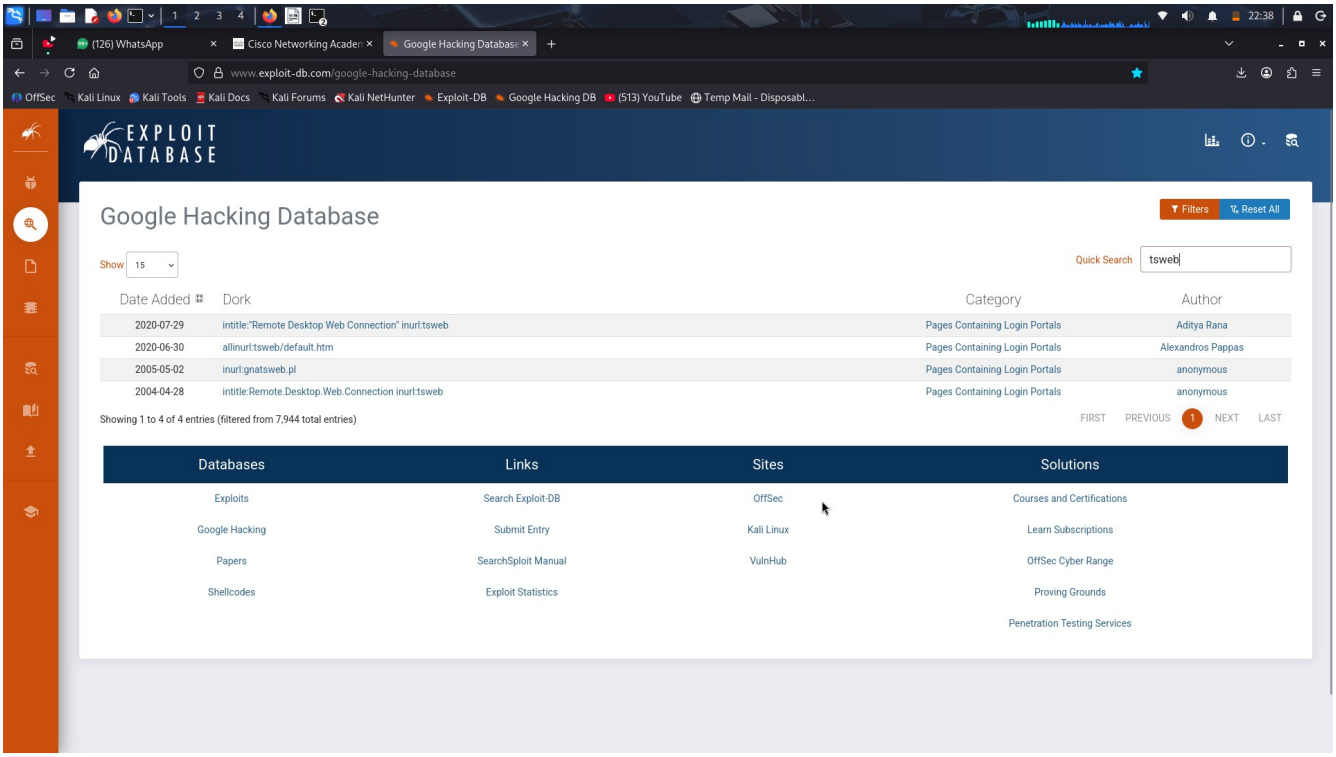
What information is provided about the Dorks?

The GHDB-ID number, the author, the date published, a brief description of the function of the dork, and a clickable link that will launch the dork in a new window.

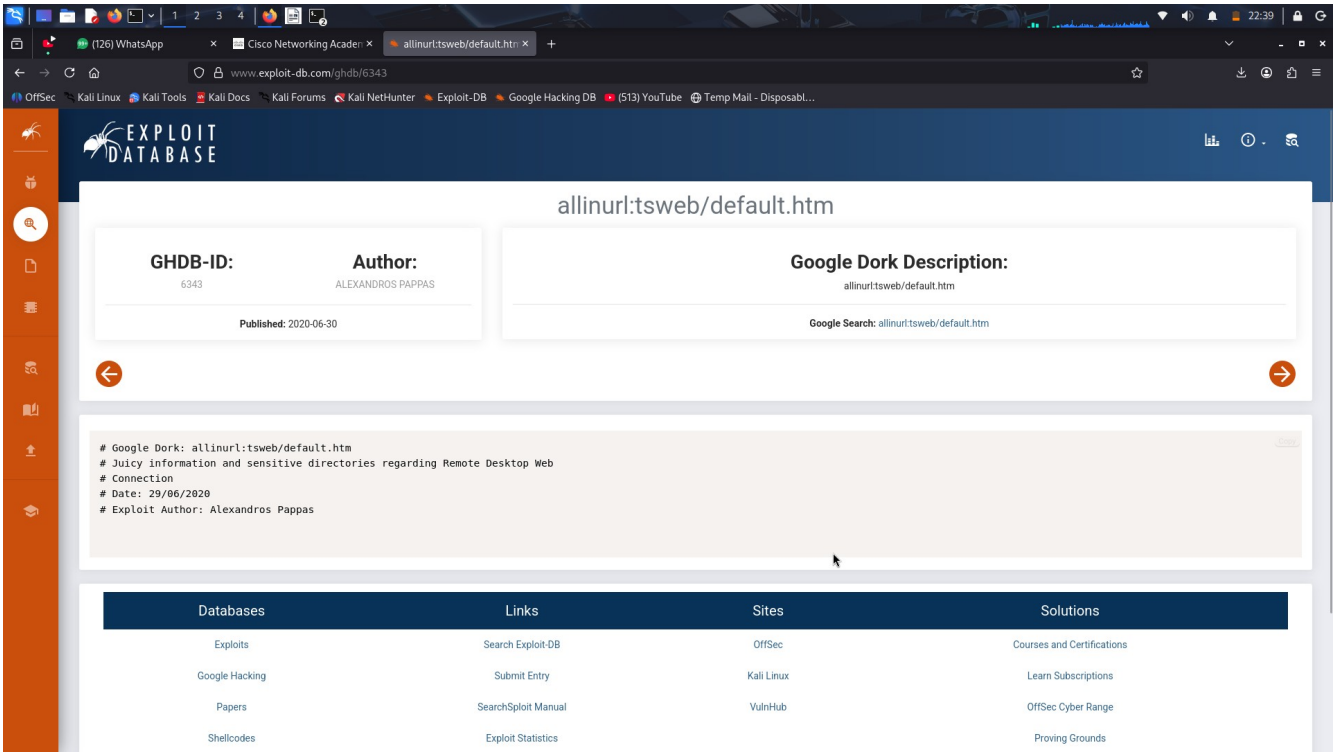


Step 3: Select Categories to find interesting Dorks.

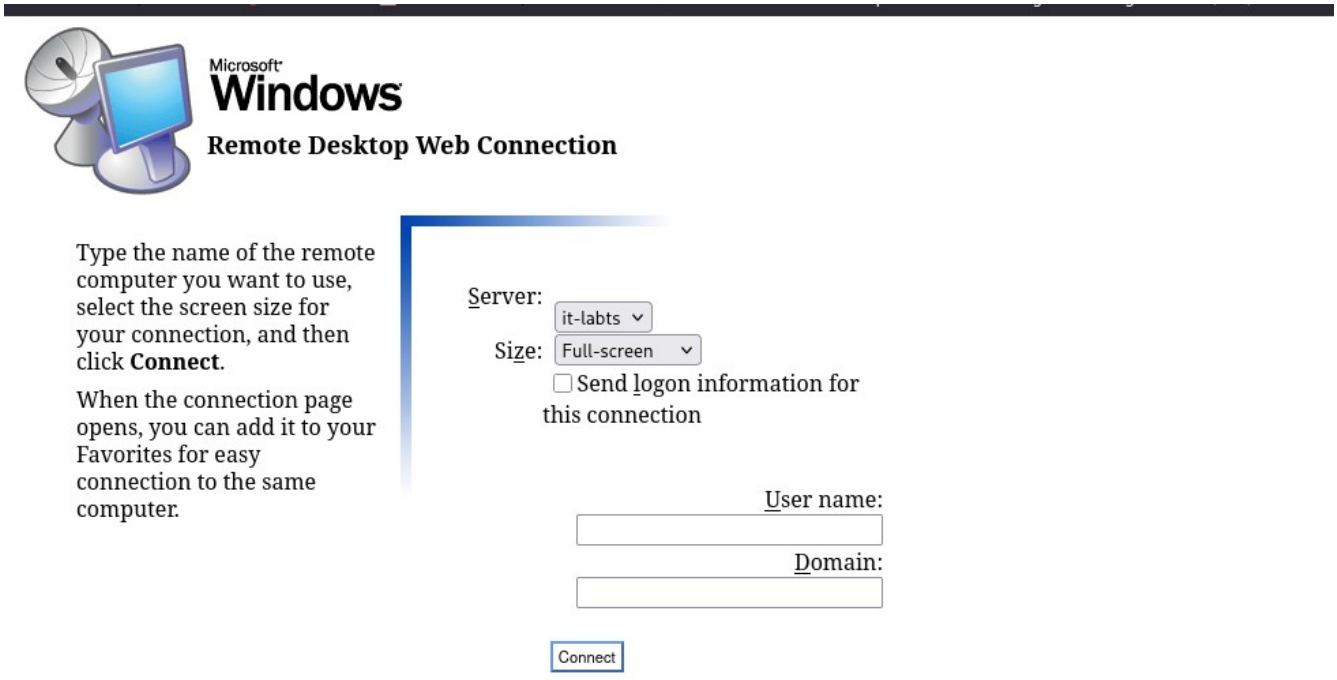
a. Conduct a search for **tsweb**.
b. Click the **allinurl:tsweb/default.htm** Dork.



Results



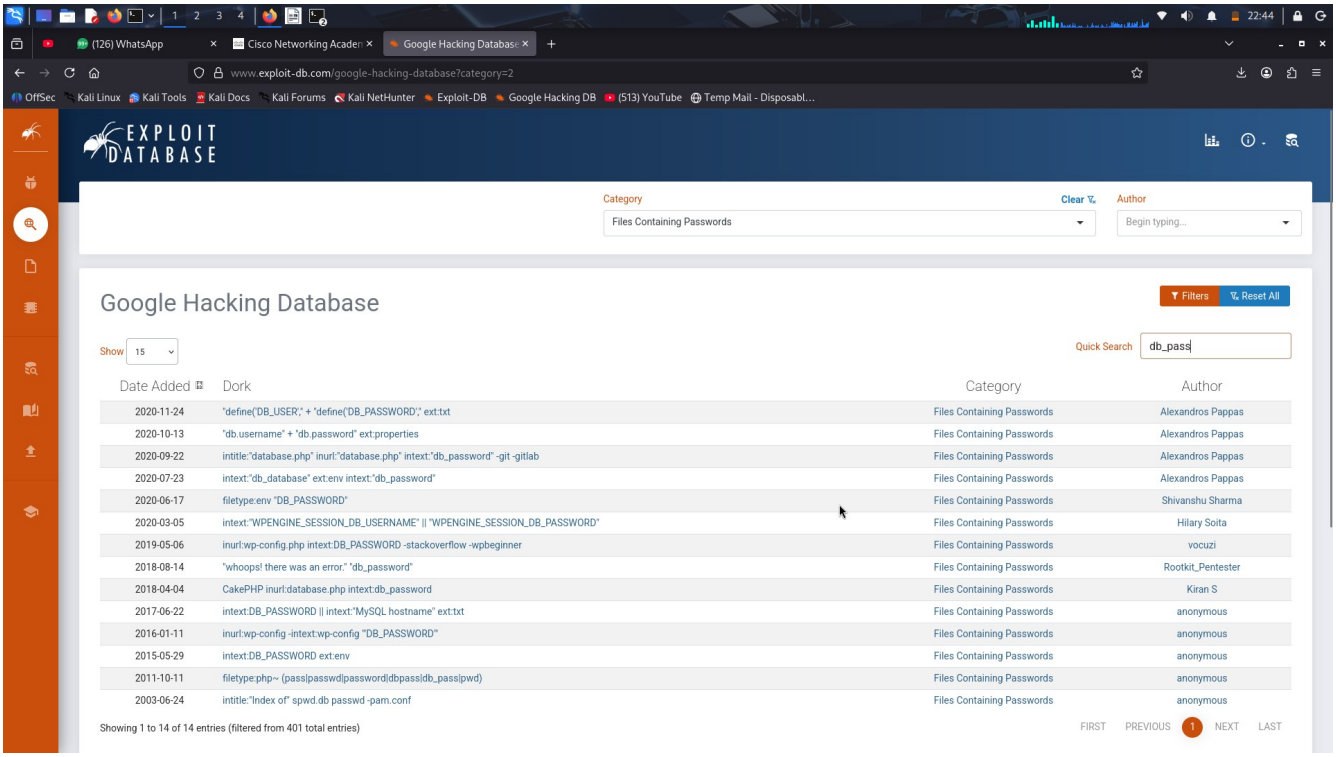
Click a few of the returned pages. Besides fields for log in credentials you may see some interesting information that could be leveraged by a hacker. For example, look at the figure. The terminal services server is running it-labts. Knowing this, a hacker can focus on the window it-labts vulnerabilities.



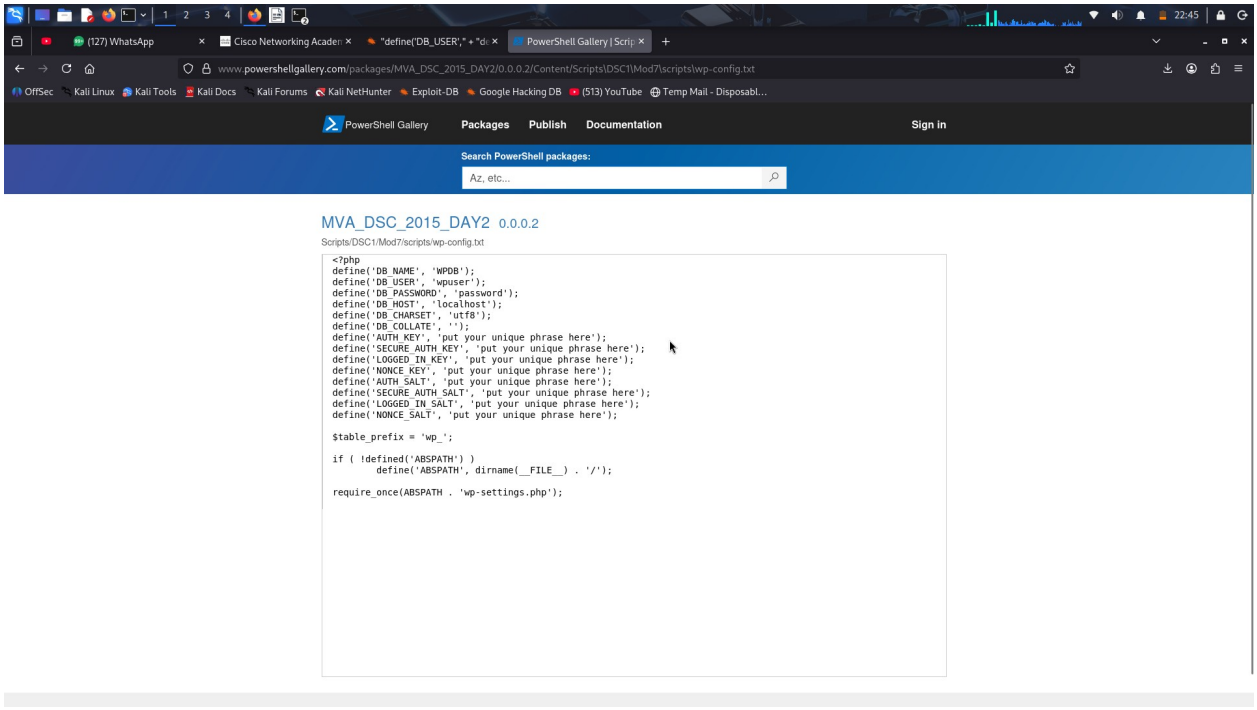
Step 4: Combine Category filters with search terms.

You can combine category filters with search terms to further refine and filter results to specific information.

a. Select **Files Containing Passwords** in the **Categories** drop down.
b. In the **Quick Search** window, type **db_pass**. This will return dork searches for database passwords.



Explore some of the search results and see what interesting information they reveal.



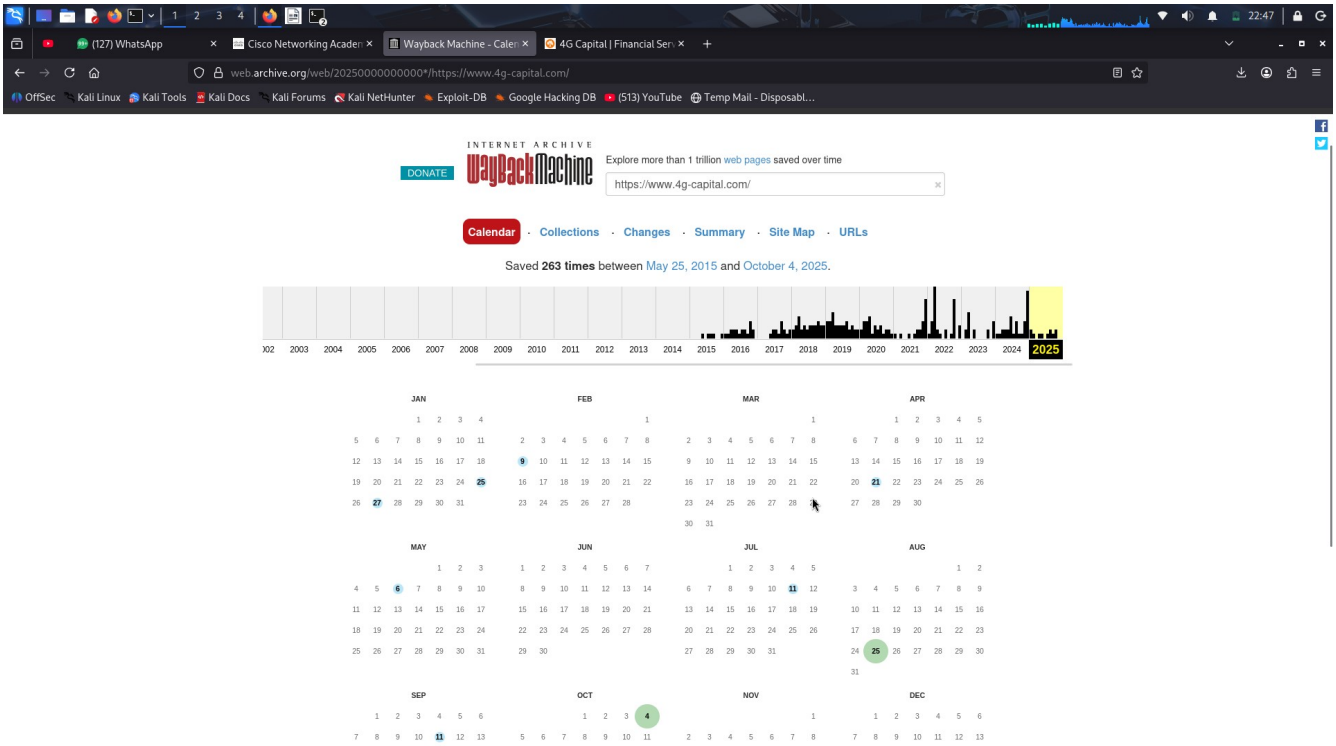
Review the course materials and try some of the searches that are shown there.

Part 3: The Wayback Machine

Website security has evolved over the decades. Websites used to publish information that is no longer considered safe. Webpage archives can reveal interesting information that is no longer available. The Wayback Machine is a useful tool for passively collecting information about a target that could be used in social engineering or other attacks. The Wayback Machine is an archive of the entire internet. It accesses every website and crawls it while taking screenshots and logging the data to a database. These endpoints can then be queried to pull down every path the site has ever crawled.

Step 1: Explore the Wayback Machine database.

a. Navigate to https://web.archive.org to bring up the Wayback Machine home page.
b. Enter the URL of a target company in the Search box.

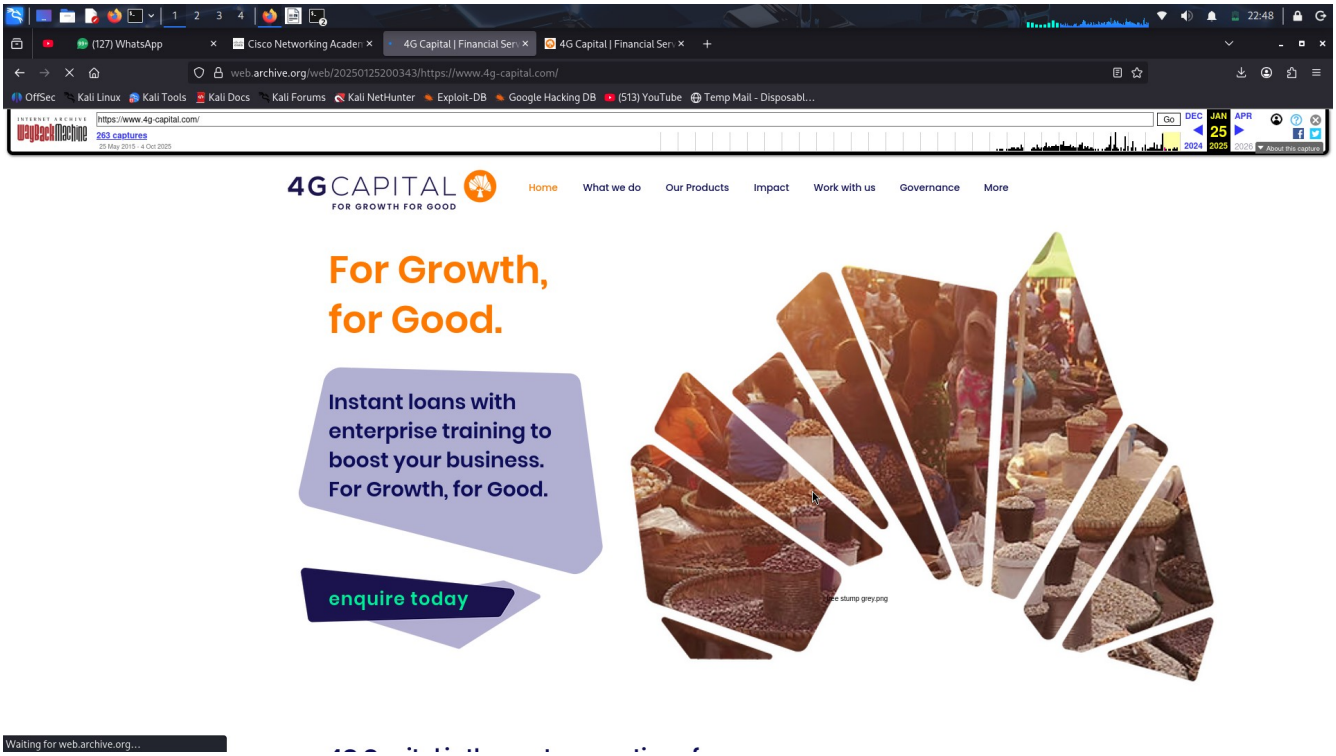


Step 2: Explore the Calendar tab.

a. Click the **Calendar** tab if not already selected.

At the top of the page is a graph that shows how many times the website has been crawled by the Wayback Machine and a calendar at the bottom showing at what day the archive entry was created. You can click these to open snapshots from the past.

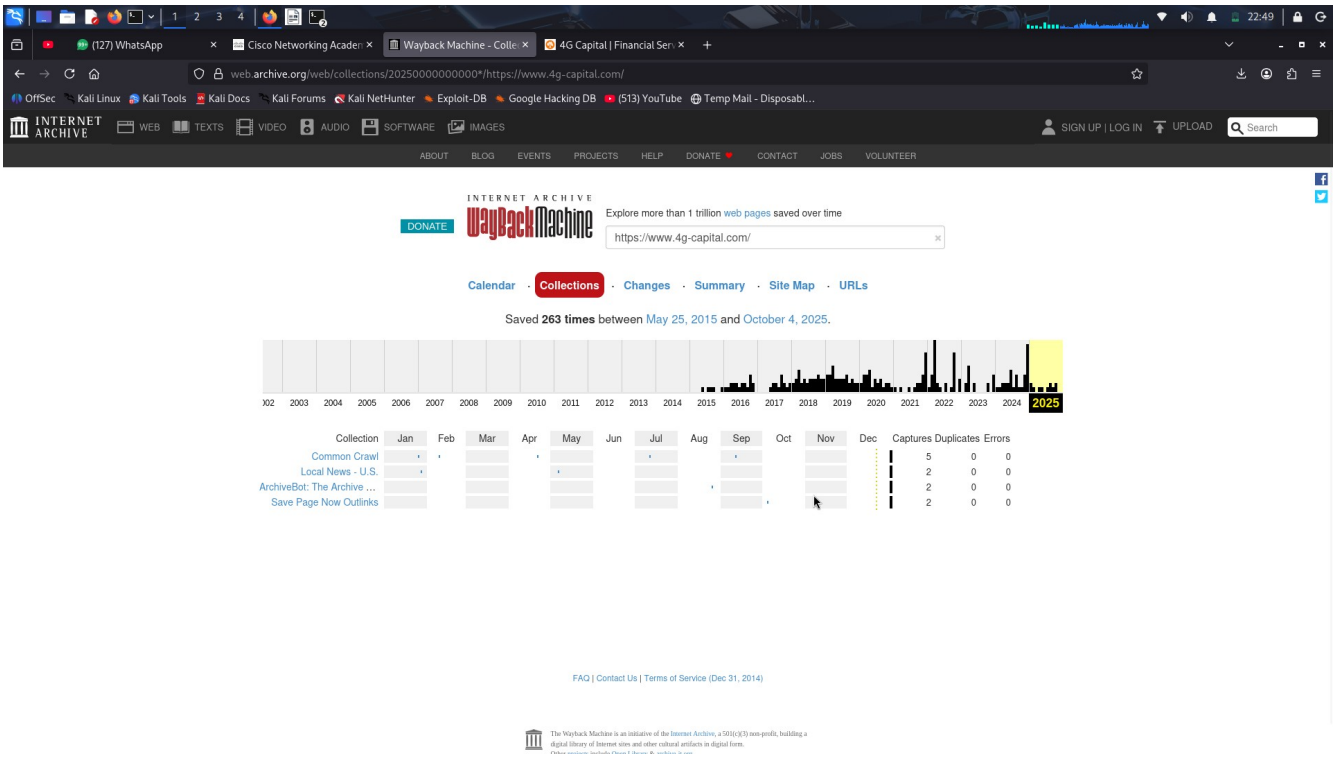
b. Select a year and a date for a snapshot in the calendar. Some dates may have more than one snapshot. Click a snapshot to open the achieved web page. Depending on the site, you may be able to navigate the page as if it was live, seeing all the dated information.



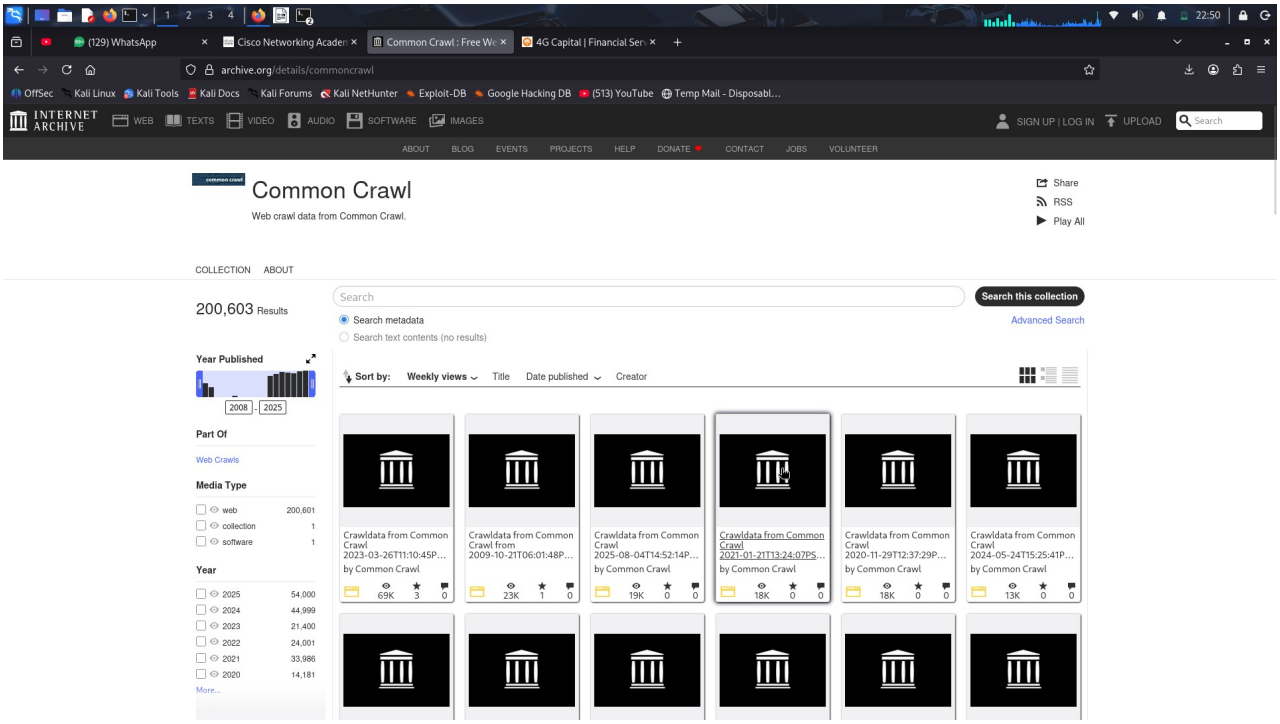
Step 3: Explore the Collections tab.

a. Click the **Collections** tab.

b. This provides archives organized by source. The collections that crawled the page are in the column on the left. The months Jan – Dec show when it was crawled over time.

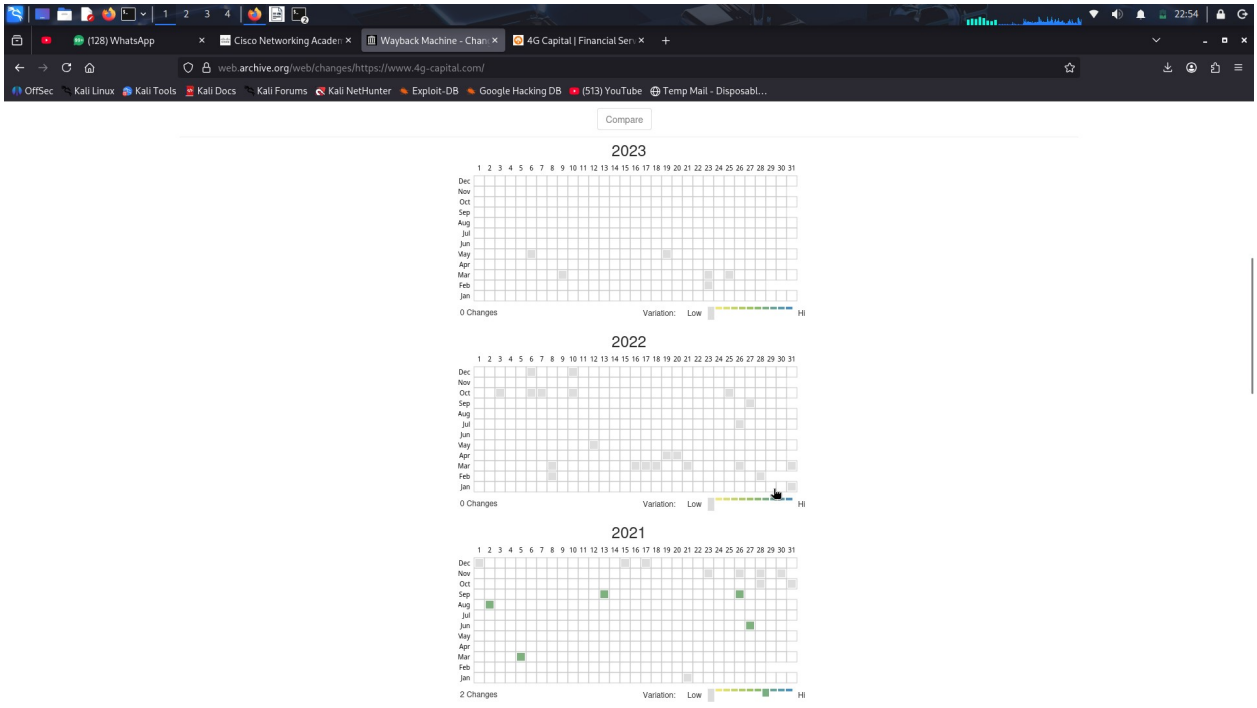


c. Click some of the collections to find out more about the collections and who runs them.



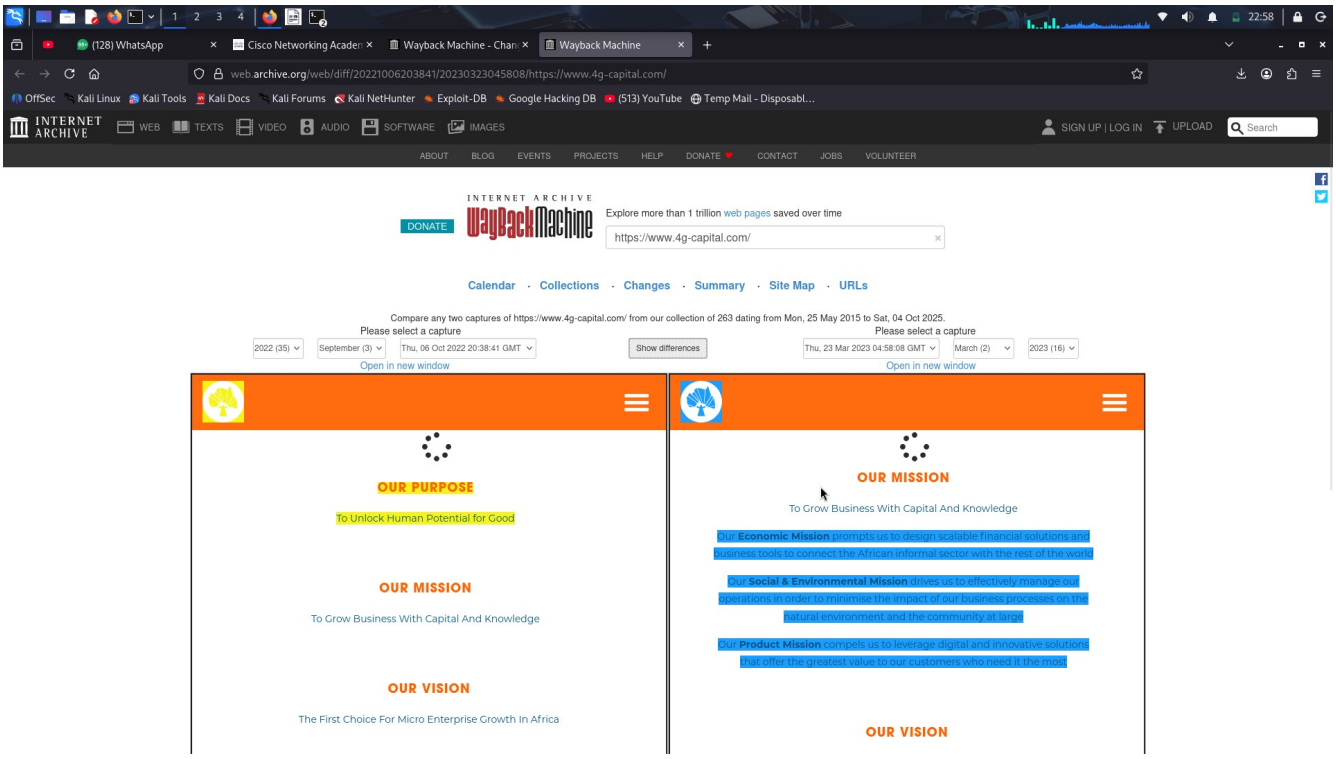
Click the **Changes** tab.

This shows how much the page has changed over time. Grey = has not changed much since the last crawl. Blue = significant changes. You can also compare changes from two captures to see what has changed.



b. Select two captures. They can be on the same day or on different days. Click the **Compare** button. Things that have changed will be highlighted.

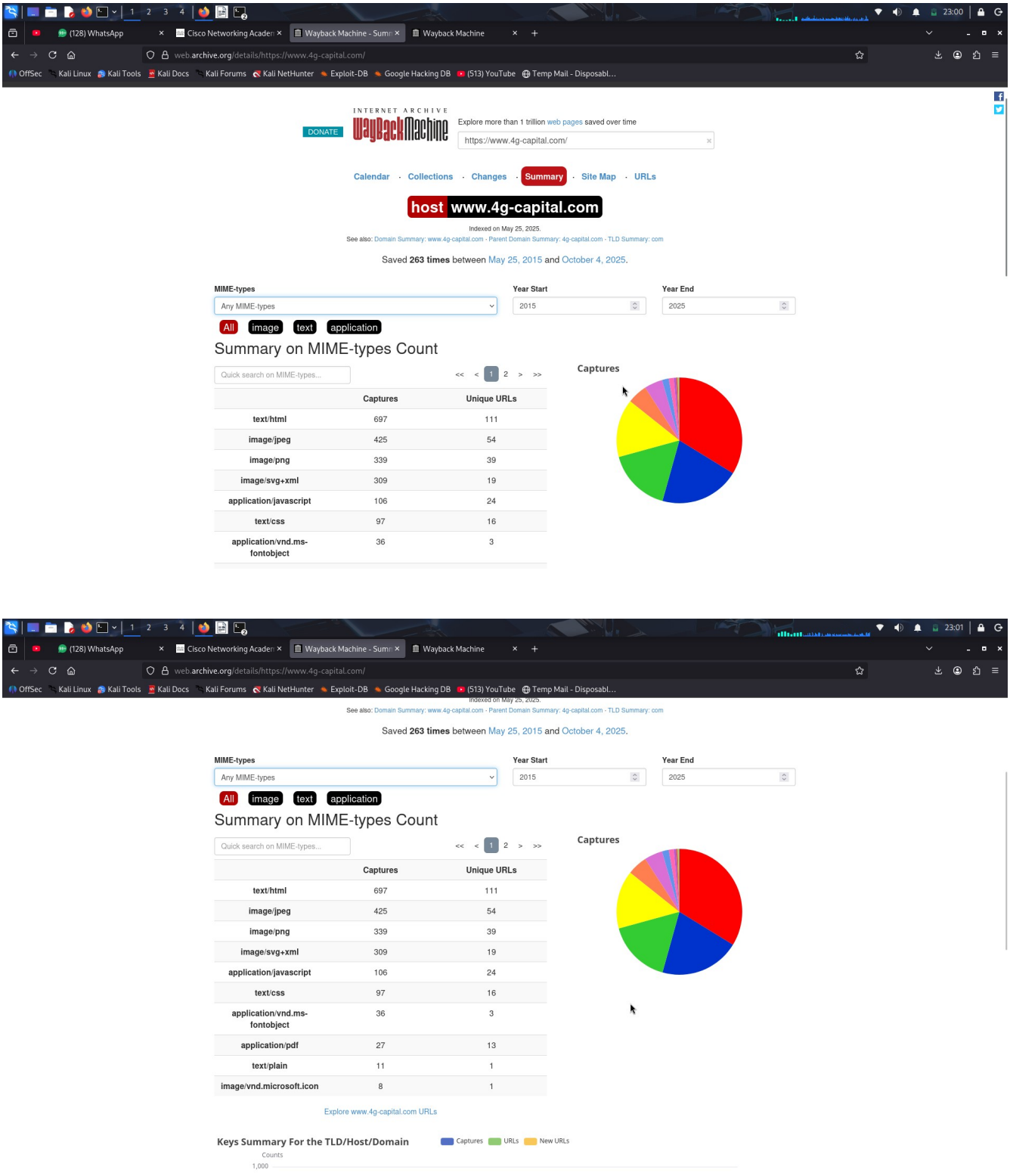
   Yellow indicates content deletion. **Blue** indicates content addition.



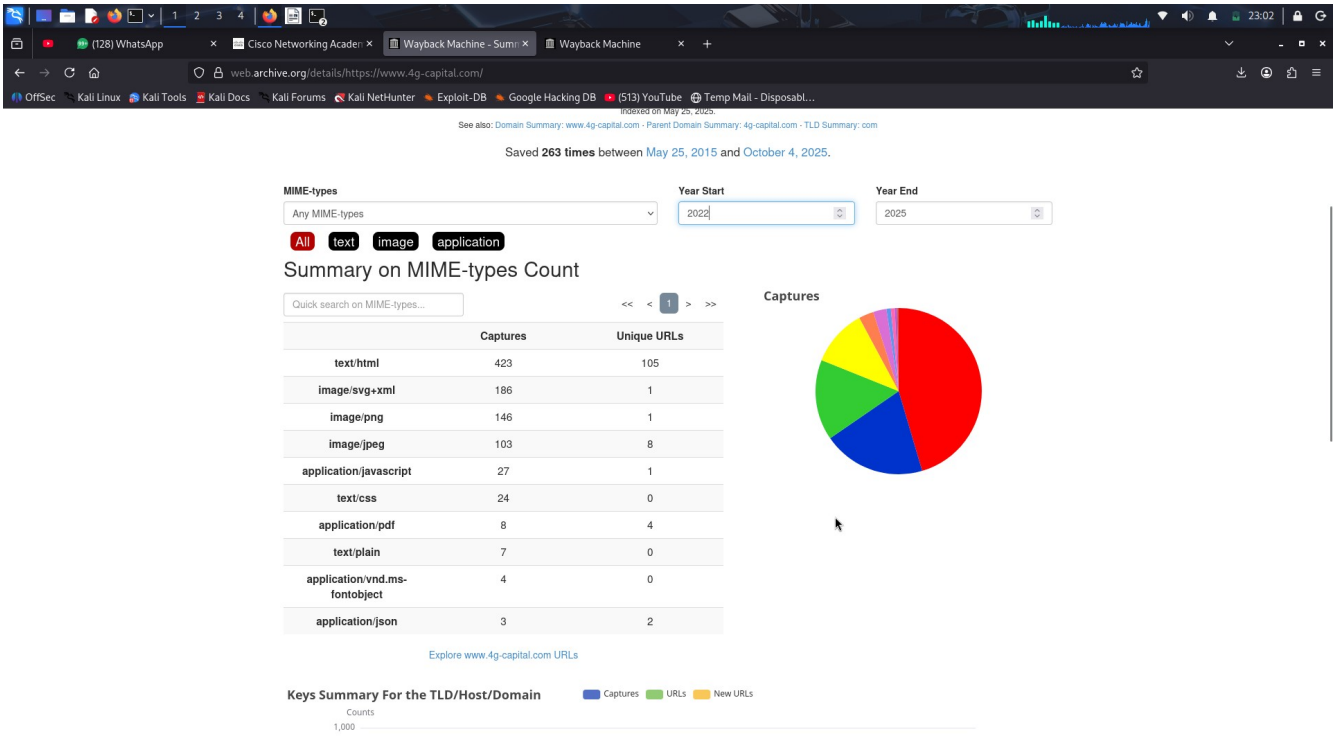
Step 5: Explore the Summary tab.

a. Click the **Summary** tab.

The summary applies to the entire domain, whereas calendar, collections, and changes are specific to the URL (single page) searched. This page shows the MIME type of the content that was hosted by the domain in the given date range. This can be text, images, javascript, etc.



b. Click the dropdown arrow in the MIME-types drop box and review the file types available.
c. Change the Year Start and Year End to see how things have changed over a time of period.

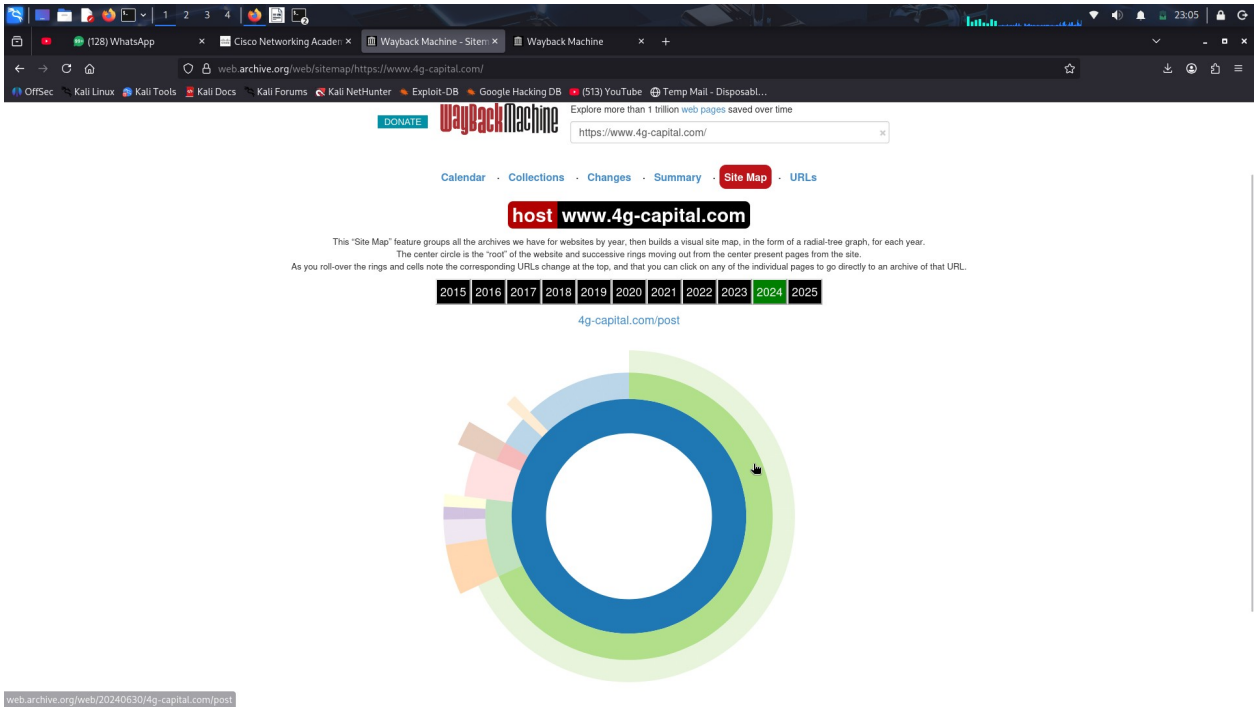


d. Click each of the data type buttons: All, text, application, image, message, audio, video, and explore the information revealed.

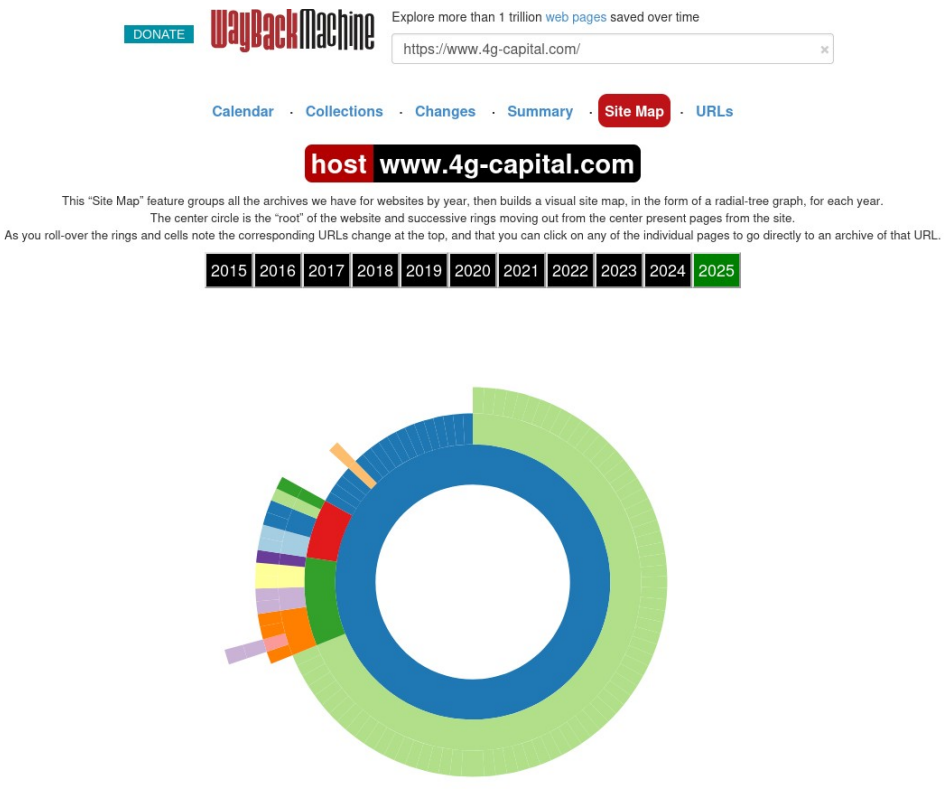
Step 6: Explore the Site Map tab.

a. Click the **Site Map** tab.

The Site Map also applies to the entire domain. The center circle is the "root" and all the rings that surround the center circle are the various pages or trees of the web site. The further out from the root, the more complex the page is.



b. Click through the years to see in the graph how the complexity of the site has changed over time.

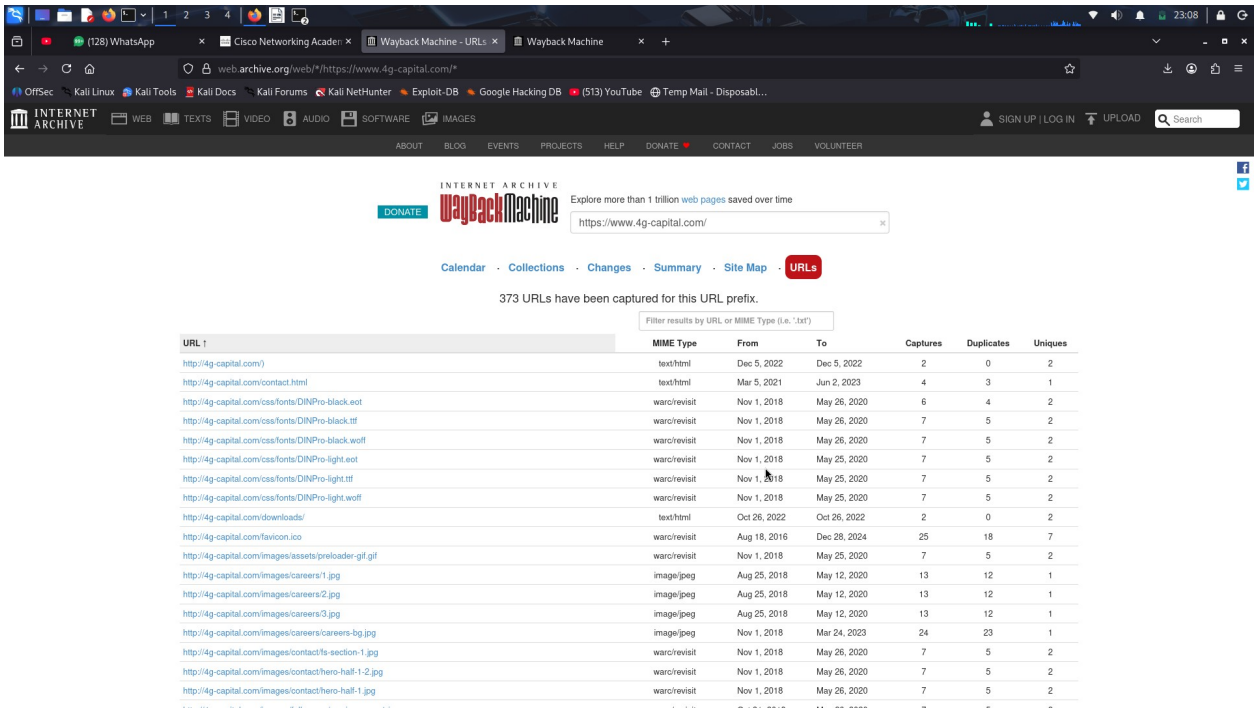


c. Click the rings and cells in the graph to open archived pages. The data in the archived pages can be used to find vulnerabilities.

Step 7: Explore the URLs tab.

a. Click the **URLs** tab.

This shows all the URLs containing the domain prefix.



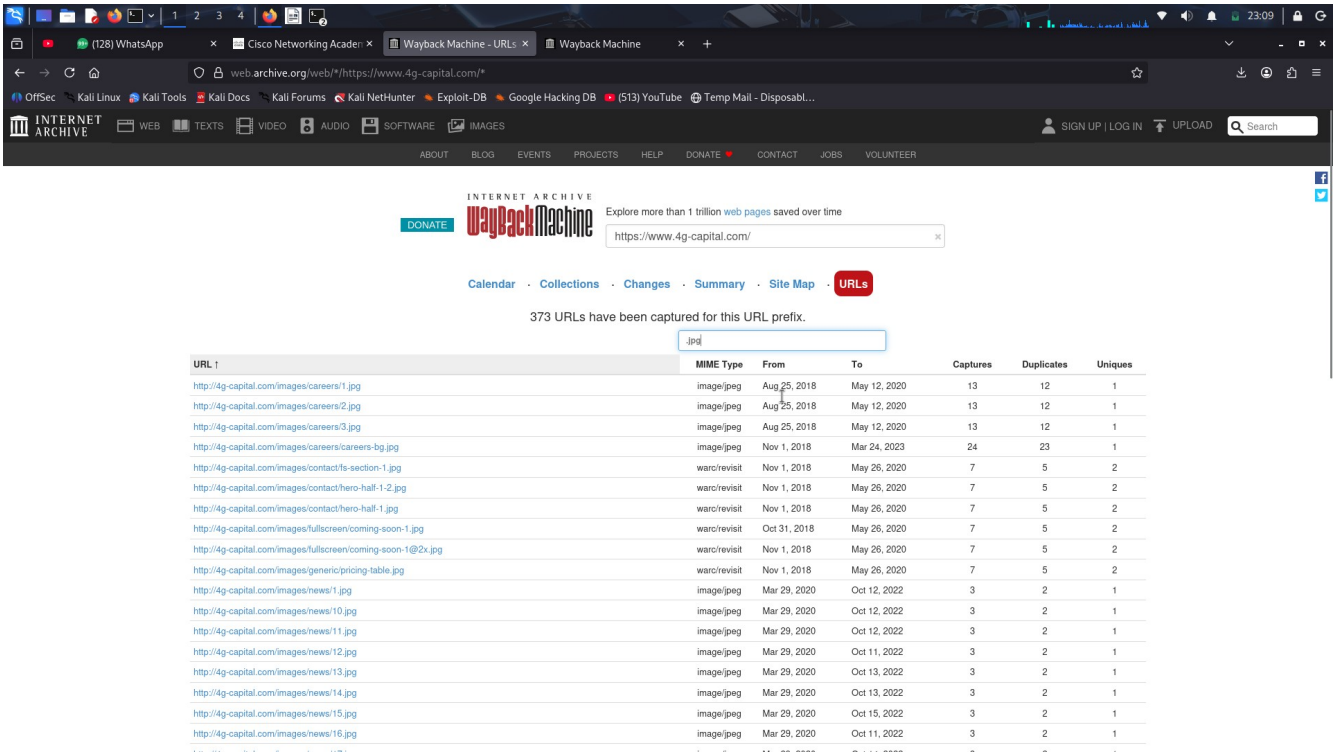
b. Use the filter box on the right of the page to search for specific files such as anything that ends in ".bak" to see if they contain any interesting backup information.

Note: Depending on the site searched, this may or may not return any content.

c. Experiment with other filters. Some may return content, some may not. Also, experiment with these filters on different domains. Some filters to try:

- .zip
- .backup
- .config
- .csv
- .pdf
- /api/
- /admin/

Not only can you find interesting files by searching the Wayback Machine archives, but careful inspection of the data can lead to finding potential vulnerabilities.



URL ↑	MIME Type	From	To	Captures	Duplicates	Uniques
http://4g-capital.com/images/careers/1.jpg	image/jpeg	Aug 25, 2018	May 12, 2020	13	12	1
http://4g-capital.com/images/careers/2.jpg	image/jpeg	Aug 25, 2018	May 12, 2020	13	12	1
http://4g-capital.com/images/careers/3.jpg	image/jpeg	Aug 25, 2018	May 12, 2020	13	12	1
http://4g-capital.com/images/careers/careers-bg.jpg	image/jpeg	Nov 1, 2018	Mar 24, 2023	24	23	1
http://4g-capital.com/images/contact/fs-section-1.jpg	warc/revisit	Nov 1, 2018	May 26, 2020	7	5	2
http://4g-capital.com/images/contact/hero-half-1-2.jpg	warc/revisit	Nov 1, 2018	May 26, 2020	7	5	2
http://4g-capital.com/images/contact/hero-half-1.jpg	warc/revisit	Nov 1, 2018	May 26, 2020	7	5	2
http://4g-capital.com/images/fullscreen/coming-soon-1.jpg	warc/revisit	Oct 31, 2018	May 26, 2020	7	5	2
http://4g-capital.com/images/fullscreen/coming-soon-1@2x.jpg	warc/revisit	Nov 1, 2018	May 26, 2020	7	5	2
http://4g-capital.com/images/generic/pricing-table.jpg	warc/revisit	Nov 1, 2018	May 26, 2020	7	5	2
http://4g-capital.com/images/news/1.jpg	image/jpeg	Mar 29, 2020	Oct 12, 2022	3	2	1
http://4g-capital.com/images/news/10.jpg	image/jpeg	Mar 29, 2020	Oct 12, 2022	3	2	1
http://4g-capital.com/images/news/11.jpg	image/jpeg	Mar 29, 2020	Oct 12, 2022	3	2	1
http://4g-capital.com/images/news/12.jpg	image/jpeg	Mar 29, 2020	Oct 11, 2022	3	2	1
http://4g-capital.com/images/news/13.jpg	image/jpeg	Mar 29, 2020	Oct 13, 2022	3	2	1
http://4g-capital.com/images/news/14.jpg	image/jpeg	Mar 29, 2020	Oct 13, 2022	3	2	1
http://4g-capital.com/images/news/15.jpg	image/jpeg	Mar 29, 2020	Oct 15, 2022	3	2	1
http://4g-capital.com/images/news/16.jpg	image/jpeg	Mar 29, 2020	Oct 11, 2022	3	2	1