Exposing the Dark Side of Google Dorks: How I Extracted Millions of Data.

Pwndec0c0

.

Follow

Published in

OSINT Team

4 min read

.

Nov 11, 2024

Listen

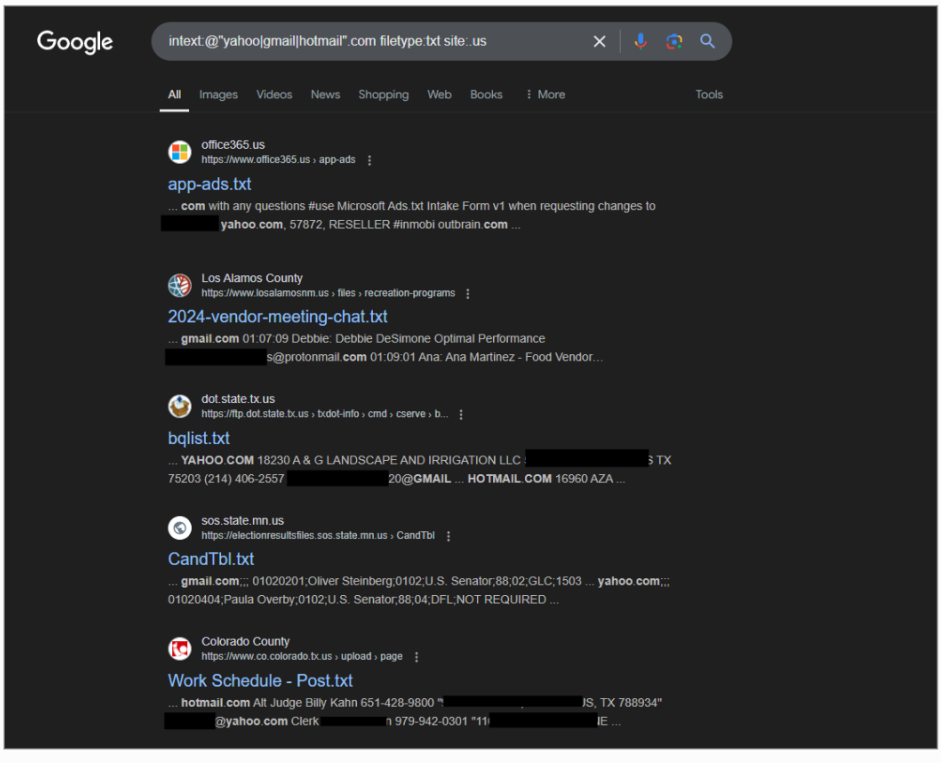
Share

More

How I was able to scrape large data of emails from the top mail domains(gmail,yahoo,hotmail etc..), .gov, .edu on of all countries and with private email as well with just using 2 simple tools **google dorks** and **bash** Let's start with the how can this be useful... Beware of my words I'll be using borrow,accidentally, not intended trust me its true 😉

Email scraping, is useful for legitimate purposes like marketing campaigns and data analysis, can also be exploited for malicious activities such as spam,phishing attacks, malware distribution.

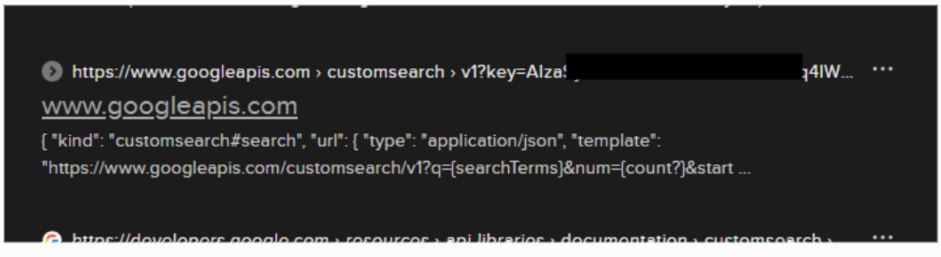
I'll show you how easily you can gather millions of emails just using google dork/ google hacking. Now what's google dork?? google it! don't be lazy but if your'e already aware of it Noiceeeee. here's a sample of the google dork I used.
intext:@"yahoo|gmail|hotmail".com filetype:txt site:.us



Results will give you sensitive information already
Now we can manually check and grab those names, emails,

addresses and other sensitive information for the good thing we're planning to use it

But come on!? "Manually??" we're going to automate this using bash for our sample; We'll be using **Google's custom search** this will require for you to create an account and generate your own API Keys. *But we're not going to do that* were just going to **"borrow"** someones API key for the sake of showing this sample, I'll be dorking this as well... and let's see if we can find one



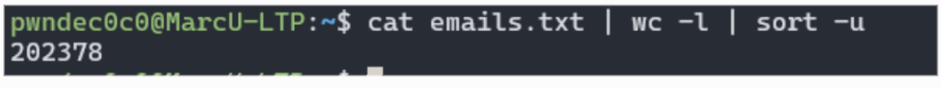
Ohhh look I accidentally found one!
now we're going to test this via curl to
curl "https://www.googleapis.com/customsearch/v1?key=[API KEY]&cx=[CX]u&q=[SEARCH STRING]"

Ohhhh it werked!
Now we’re going to create our script to gather emails or any data from google.
#Search String
intext:@"yahoo|gmail|hotmail".com filetype:txt site:.us
#we are checking for any webpages for the pattern above on google#!/bin/bash
#we are getting the number of google pages results here
PAGECNT=$(curl -s "https://www.googleapis.com/customsearch/v1?q=[SEARCH STRING]&key=[API KEY]&cx=[cx]&start=1" | jq -r '.queries.nextPage[0].startIndex')
#display total number
echo "$PAGECOUNT"

#we're going to iterate in those pages to gather all URL results and save it
for i in (seq 1 $PAGECNT)
do
curl -s 'https://www.googleapis.com/customsearch/v1?q=[SEARCH STRING]&num=10&key=[API KEY]&cx=[CX]&start='$i | jq -r '.items[].link' | anew links.txt
done

#we're going to iterate on those URLs and check for email patterns using regex
for link in `cat links.txt`;do
curl -s $link |grep -E -o '[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}' | tee -a emails.txt
done

Guess what!? we're already able to fetch these emails



we were able to gather 200k email with just using google dorks and bash
Now as you can see we're just limiting our test on text files we can expand this to other filetypes as well including csv,log files, back up files etc this will give you more data, and since the site focuses on ".us" only this will isolate our scrapping of emails with only to domains that contains ".us" we can also change this to expand our data gathering you can also set this for specific target domains to check for any leaks.
As a Penetration Tester this is a great tool to have specially if your doing blackbox testing you can isolate to the targets domain and bruteforce those accounts for any weak passwords. You can improve the script that I showed you

fine tune it, run it on thread, check for emails validity etc.. customize it to help you on your use case Just remember do it ethically ;).

Take Aways

alway use “**Plus addressing**” when using your email to register,subscribe or anything related to you giving your email. E.g subscribing to Netflix use “youremail1337+netflix.com” no worries you can still receive your email in that way, advantage of this is once the email you received that was intended to “youremail1337+netflix.com” does not come from Netflix you are already aware that your email from Netflix was leaked or sold to third parties(I’m not saying Netflix sells your data! come on guys! or is it? LOL) anyways hope this help :) Chow! Thanks,

▬▬▬▬▬▬▬▬✨ Support me ✨▬▬▬▬▬▬▬▬

Follow me on X : https://x.com/pwndec0c0
Subscribe to my YT : https://www.youtube.com/@pwndec0c0Tv
Buy me a coffee :