&cyber data

“From bits to information”

Open Source Intelligence

Outline

- Threat Hunters.
- Open Source Data.
- OAuth 2.0, APIs and Web scraping.
- Photographic sites.
- Social Networks.
- Crowd—sourced content.
- Google hacks, Maltego and Wayback.
- Shodan

Facebook – log in or sign up
https://www.facebook.com
facebook
Facebook helps you connect and share with the
people in your life.

&cyber data

cyber & data

"From bits to information"

Threat Hunters

Threat Hunters

- **Collect and process data**. This is a continual process and involves collecting data from internal systems - such as from SIEM generated data - and also from external sources.
- **Establish a hypothesis**. This defines the basic reason for the hunting and will match to a business-oriented goal and which is relevant to the company. For example, a hypothesis could be that "A known threat actor could pay an insider to get access to our company and plant a backdoor Trojan on the network."
- **Hunt**. This then backs up the hypothesis, and the hunter will go and try to find the required information to either corroborate or dismiss the hypothesis.
- **Identify threats**. The proving of a hypothesis will then lead to a possible threat being identified. An organisation must then understand the scope of this threat, and whether it is real or not.
- **Respond**. If a real threat is identified, it is important to create the required response and to put in place plans to further investigate, or create mitigation plans.

&cyber data

&cyber data

“From bits to information”

Open Source Data

Open Source Intelligence

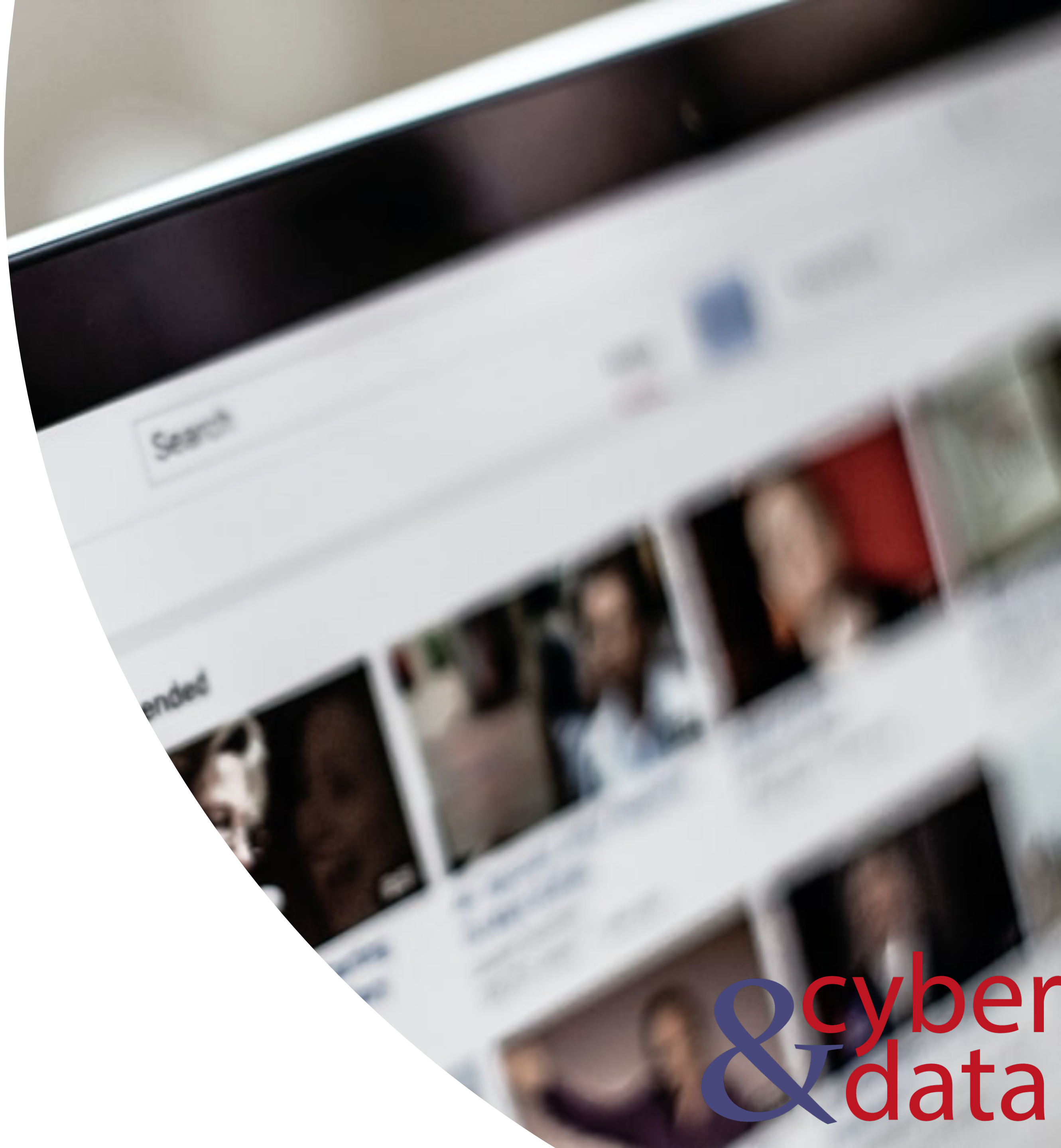
- **Social networks.** This includes sources such as Facebook and Twitter and where individuals consent for their data to be seen either in a public space (such as with Twitter) or for their trusted contacts (as with Facebook).
- **Crowd-sourced content.** Within crowd-sourced sites, groups of people come together to debate topics and/or answer questions. Typical crowd-source sites include Reddit and Stack Overflow. In fact, Reddit is now one of the top Web sites in the world.
- **Wikis.** This includes sites which have been shared as Wikis. These can include publicly sourced information from sites such as Wikipedia, or less publicly-facing Wikis, such as FANDOM. The open-source information could contain information around company locations, key company employees, and links to sensitive company publications. The problem with less publicly available Wikis is that they are often created for sharing within a given project, but are still publicly scannable.
- **Google Hacks (aka Google Dorking).** This involves using extensions within the Google search engine, and often target sites and document types.

Open Source Intelligence

- Organisational web site. A typical source of intelligence might be from an organisation's Web site such as for key staff, organisational structure, locations, telephone numbers, and so on.
- Job sites. Job postings can reveal information around the technologies that are used within an organisation, along with other information on the structure of the organisation and their locations. For example, a posting for an Oracle database designer might thus reveal some information around the type of data architecture that the company is using.
- Business databases. This includes government sites which publish details of companies. One example of this in the US is the EDGAR database, and which reveals information about publicly trading companies.
- Photographic sites. One of the largest growth areas within on-line content is in the hosting of photographs from users. Instagram, Flickr and Pinterest are three key sites which are popular for uploading photographs and short videos. Typically the location information within the EXIF data is stripped from the content.

Open Source Intelligence

- Music information, services such as Spotify support the publishing of playlists to friends.
- Video. The viewing of video information and likes/dislikes are normally kept private, but the posting of comments on videos, such as within YouTube, may reveal information into an open environment.
- Interest sites. Sites of special interest are often good sources of open source intelligence, as users typically use a pseudo-ID and which can be linked to their core identity. The sites, too, are normally moderated for fake users, and for those who are abusing the terms and conditions of the site. LinkedIn is one site which requires users to define their proper name and often to provide their job history. Those who abuse the terms and conditions of the site are likely to be spotted, and removed from the site.
- Location-based services. For some, the natural extension of sharing information on social media is to share location information on sites such as Foursquare. In a public space, it is typically associated with either a lifelogging - and where users document their life to others from their travels - or for swarm activities, and where users aim to swarm together. A swarming activity might relate to a political demonstration, and where activists aim to find other activists.



&cyber
data

Open Source Intelligence

- Influence. There are a number of sites which aim to measure the impact of things like users, brands and organisations. They typically gather federated access for a key social media account related to the user, and then measure their impact. A popular site is Kred and which publishes impact metrics, such as the number of followers that a user has, and the number of retweets that they gain. These metrics are then used by other sites, in order to publish league tables of key influencers.

- Blogs. These sites include Tumblr and Medium, and are often sources of opinion pieces.

- Collaboration. These types of sites support collaboration, and they typically constrain the open data to a range of trusted users. A typical service is Doodle.

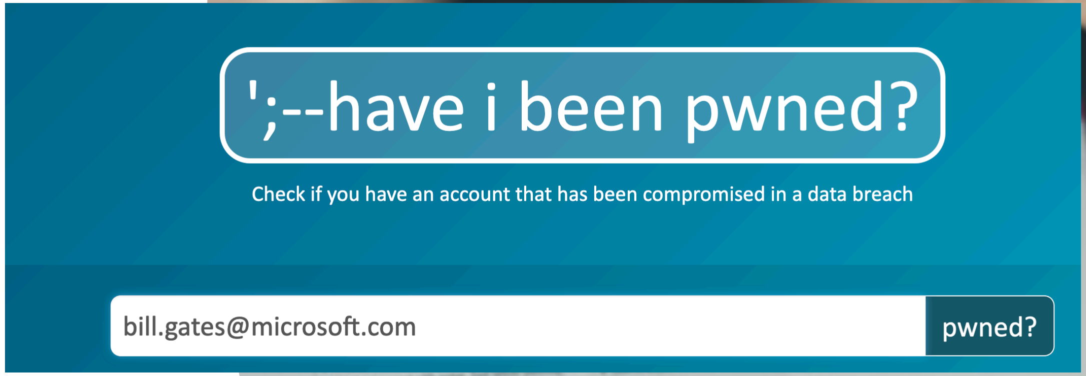
- Dark Web. While not quite open source, as it is not possible to access the Web infrastructure on the Dark Web, we can use the Tor protocol to connect to a Web server in the Dark Web, and scan for content. While the Tor project is useful for protecting privacy, the Dark Web has often been used for malicious purposes. A Tor browser can be used to connect to the open Web, and where the IP address which is defining in the log of the Web accesses will be defined with the IP address of the exit node of the Tor network. A full end-to-end encrypted connection can also be implemented onto a .onion site. It should be noted that a .onion site is not an automatic identifier that a site could have criminal activities.

&cyber data

“From bits to information”

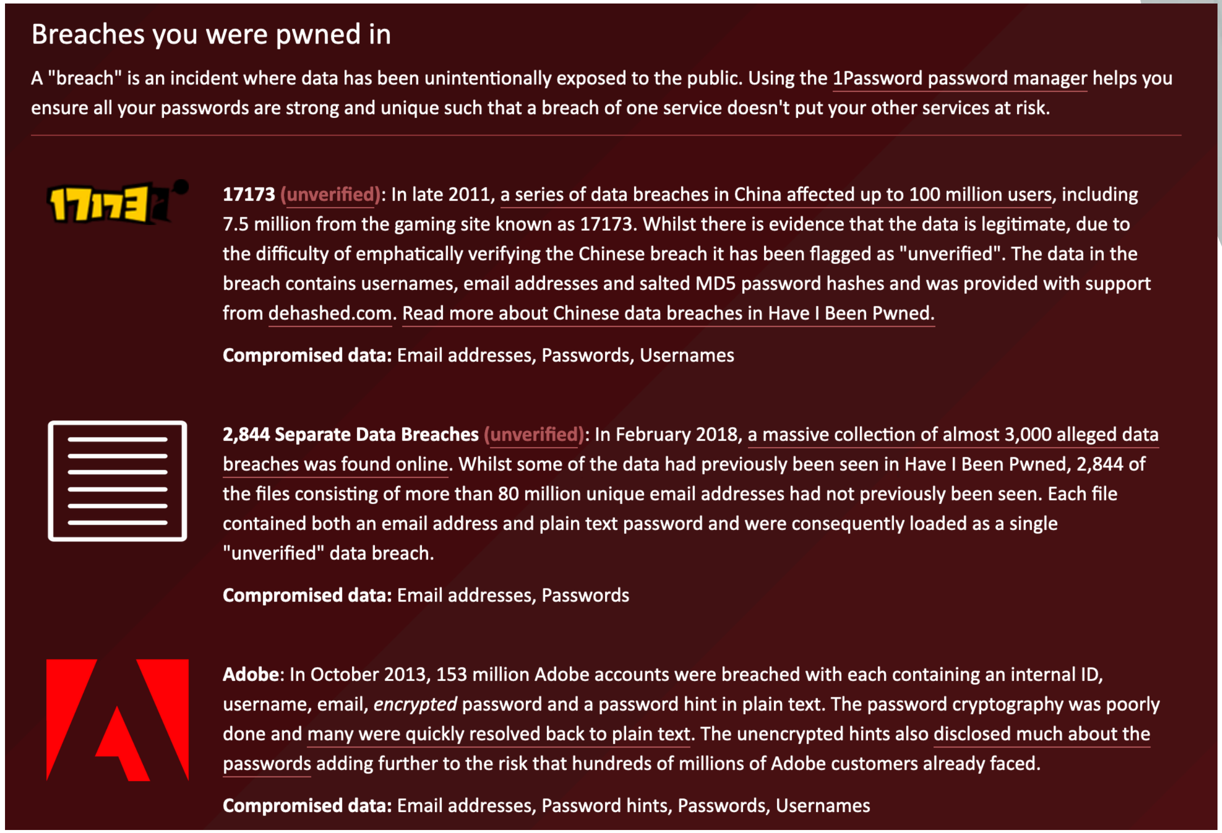
Open Source Search Data

HaveIBeenPwned



';--have i been pwned?

Check if you have an account that has been compromised in a data breach



Breaches you were pwned in

A "breach" is an incident where data has been unintentionally exposed to the public. Using the 1Password password manager helps you ensure all your passwords are strong and unique such that a breach of one service doesn't put your other services at risk.

17173 (unverified): In late 2011, a series of data breaches in China affected up to 100 million users, including 7.5 million from the gaming site known as 17173. Whilst there is evidence that the data is legitimate, due to the difficulty of emphatically verifying the Chinese breach it has been flagged as "unverified". The data in the breach contains usernames, email addresses and salted MD5 password hashes and was provided with support from dehashed.com. Read more about Chinese data breaches in Have I Been Pwned.

Compromised data: Email addresses, Passwords, Usernames

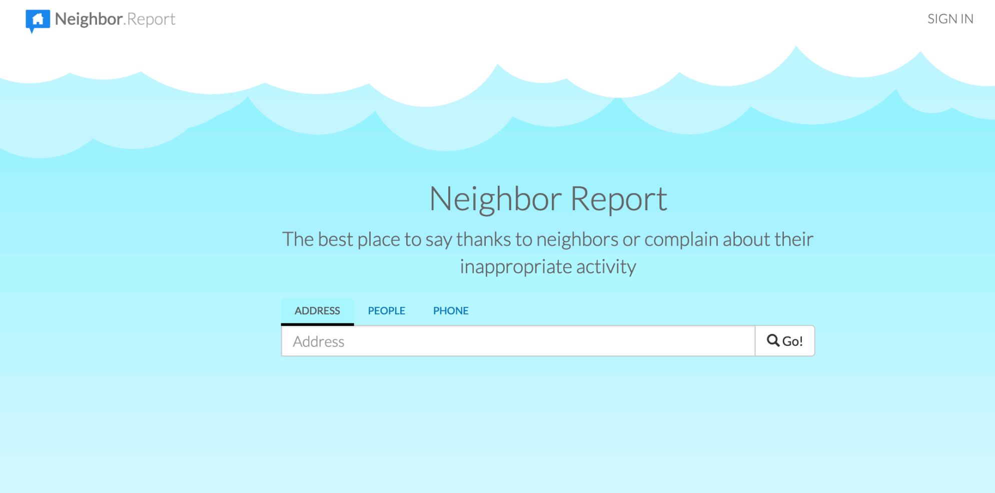
2,844 Separate Data Breaches (unverified): In February 2018, a massive collection of almost 3,000 alleged data breaches was found online. Whilst some of the data had previously been seen in Have I Been Pwned, 2,844 of the files consisting of more than 80 million unique email addresses had not previously been seen. Each file contained both an email address and plain text password and were consequently loaded as a single "unverified" data breach.

Compromised data: Email addresses, Passwords

Adobe: In October 2013, 153 million Adobe accounts were breached with each containing an internal ID, username, email, *encrypted* password and a password hint in plain text. The password cryptography was poorly done and many were quickly resolved back to plain text. The unencrypted hints also disclosed much about the passwords adding further to the risk that hundreds of millions of Adobe customers already faced.

Compromised data: Email addresses, Password hints, Passwords, Usernames

OSINT Framework



Neighbor.Report

SIGN IN

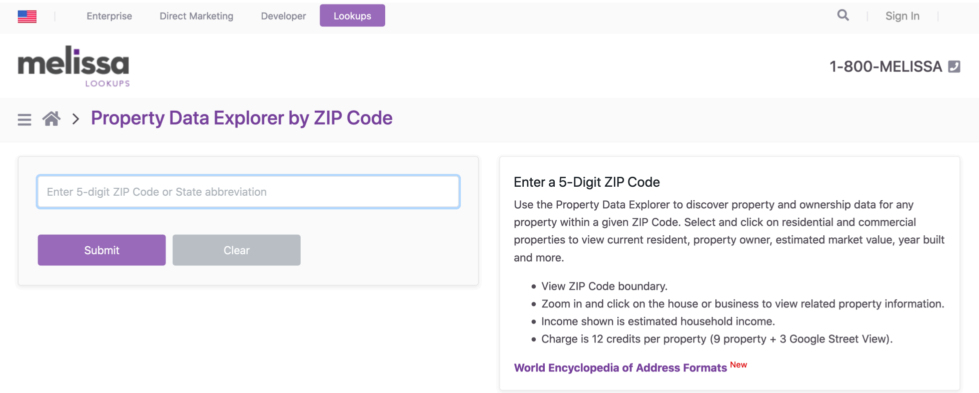
Neighbor Report

The best place to say thanks to neighbors or complain about their inappropriate activity

ADDRESS PEOPLE PHONE

Address

Go!



Enterprise Direct Marketing Developer Lookups Sign In

melissa LOOKUPS

1-800-MELISSA

Property Data Explorer by ZIP Code

Enter 5-digit ZIP Code or State abbreviation

Submit Clear

Enter a 5-Digit ZIP Code

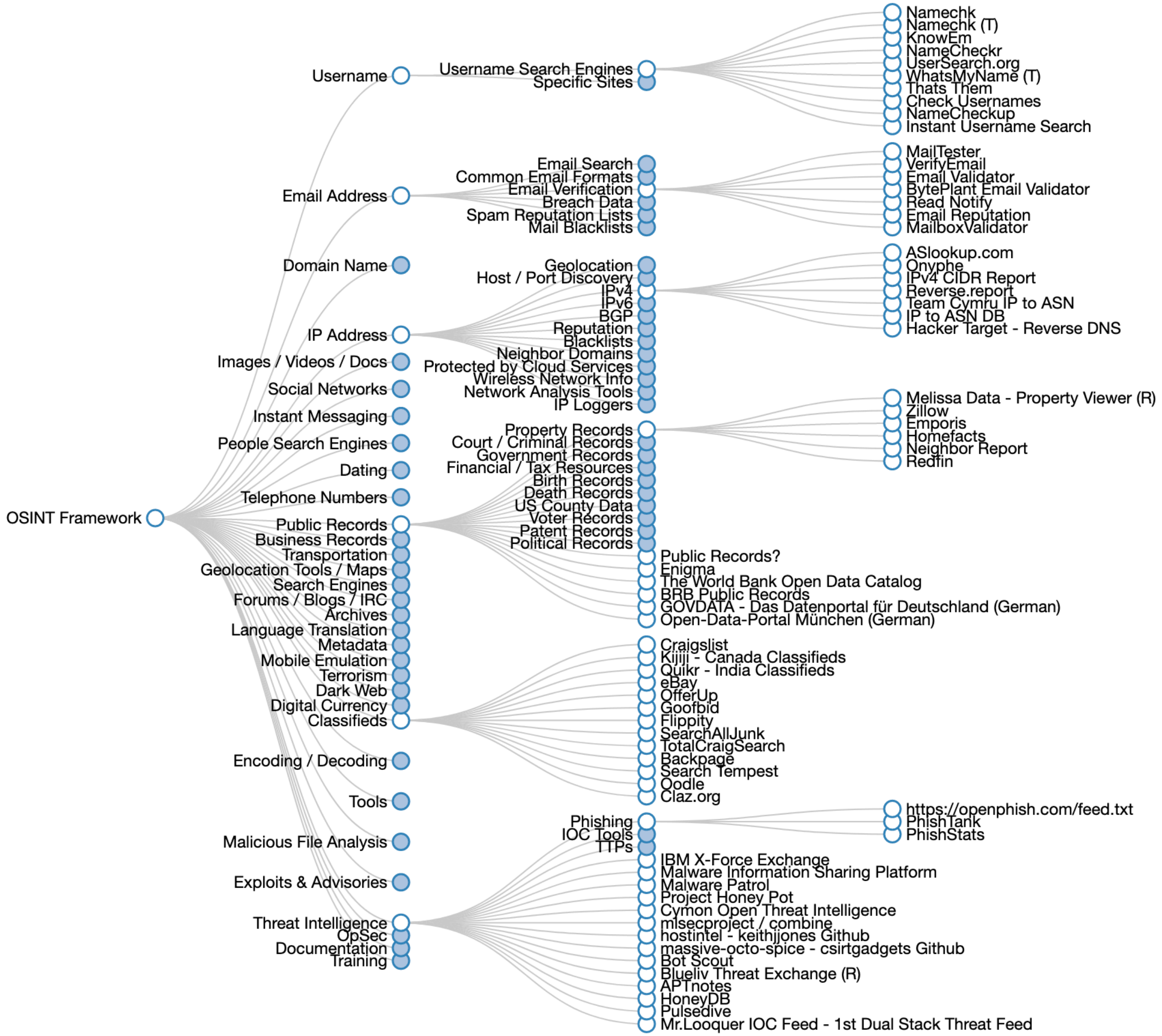
Use the Property Data Explorer to discover property and ownership data for any property within a given ZIP Code. Select and click on residential and commercial properties to view current resident, property owner, estimated market value, year built and more.

- View ZIP Code boundary.
- Zoom in and click on the house or business to view related property information.
- Income shown is estimated household income.
- Charge is 12 credits per property (9 property + 3 Google Street View).

World Encyclopedia of Address Formats New

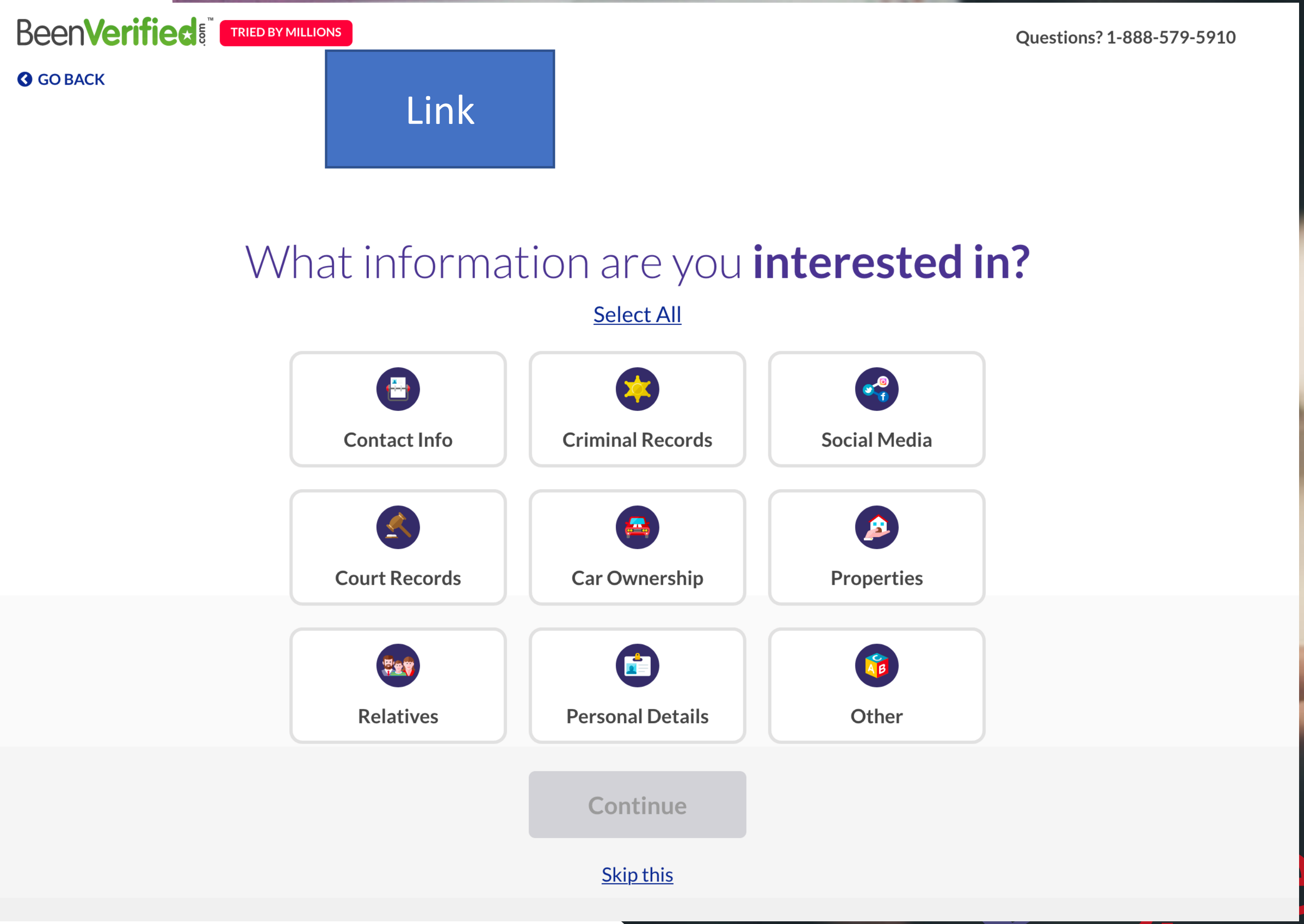
Link

OSINT Framework



BeenVerified

- **UNLIMITED** Background reports
- **UNLIMITED** Contact information
- **UNLIMITED** Phone lookups
- **UNLIMITED** Email lookups
- **UNLIMITED** Address lookups
- **UNLIMITED** Criminal records



BeenVerified.com TRIED BY MILLIONS

Questions? 1-888-579-5910

GO BACK

Link

What information are you **interested in?**

Select All

cyber & data

"From bits to information"

Photographic Sites

Photo information

- Photographic sites can be used to build up timelines of activity, or at least build up a picture of the life of someone. In 2012, a hacker known as "w0rmer," broke into a number of law enforcement sites. He then posted a photograph on Twitter of a woman holding a sign taunting investigators. Unfortunately for him, the photograph contained EXIF metadata, and which included the GPS data of the location where the photograph was taken. Investigators traced it down to a house in Wantirna South, Australia. Increasingly, though, the GPS coordinates are stripped off online pictures by the service provider, but where the serial number of the camera is still defined. A sample of the metadata contained within a photograph is:

Photo information

- Exif Image Size: 200x200
- Make: samsung
- Camera Model Name: M-G965F
- Orientation: Horizontal (normal)
- Modify Date: 2019:07:27 10:08:09
- Y Cb Cr Positioning: Centered
- Exposure Time: 1/100
- F Number: 2.40
- Exposure Program: Program AE
- ISO: 125
- Exif Version: 0220
- Date/Time Original: 2019:07:27 10:08:09

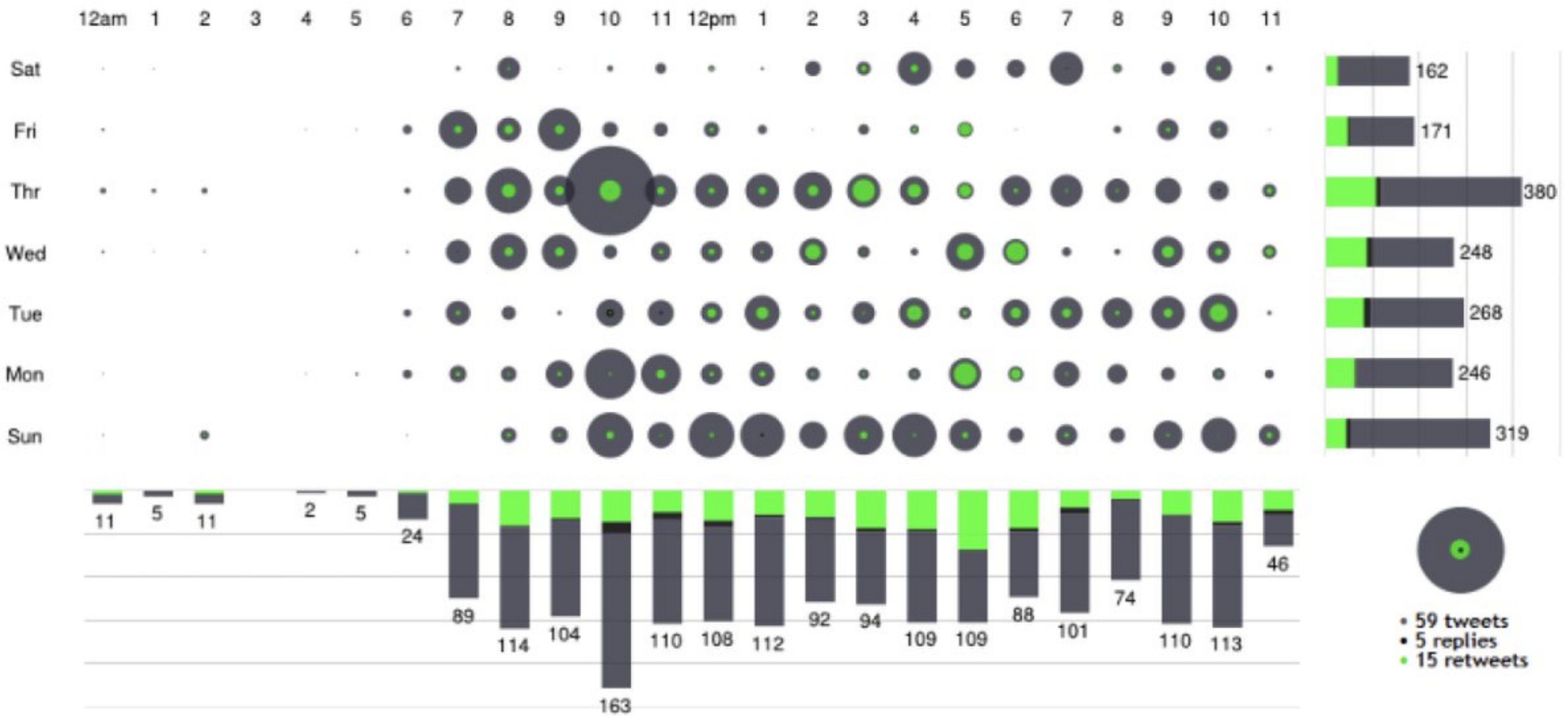
- Create Date   2019:07:27 10:08:09
- Aperture Value: 2.39
- Brightness Value: 3.58
- Image Size: 512×288
- Software: G965FXXU5CSF2
- Shutter Speed Value: 1/100
- Max Aperture Value: 1.5
- Focal Length: 4.3 mm
- Image Unique ID: I12LLKF00SM
- Compression: JPEG (old-style)
- Resolution: 72 pixels/inch
- Thumbnail Length: 13,528

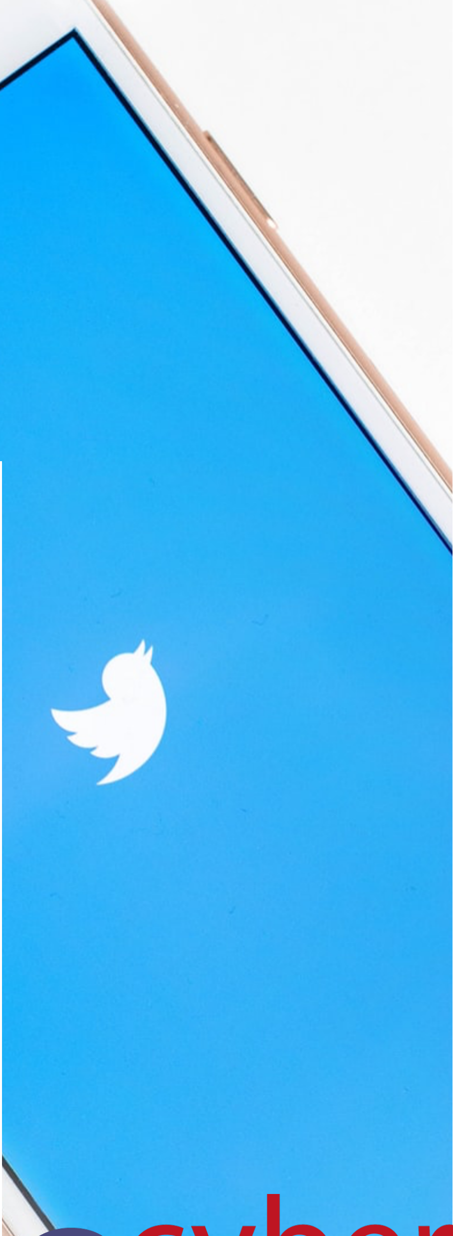
&cyber data

"From bits to information"

Social Networks

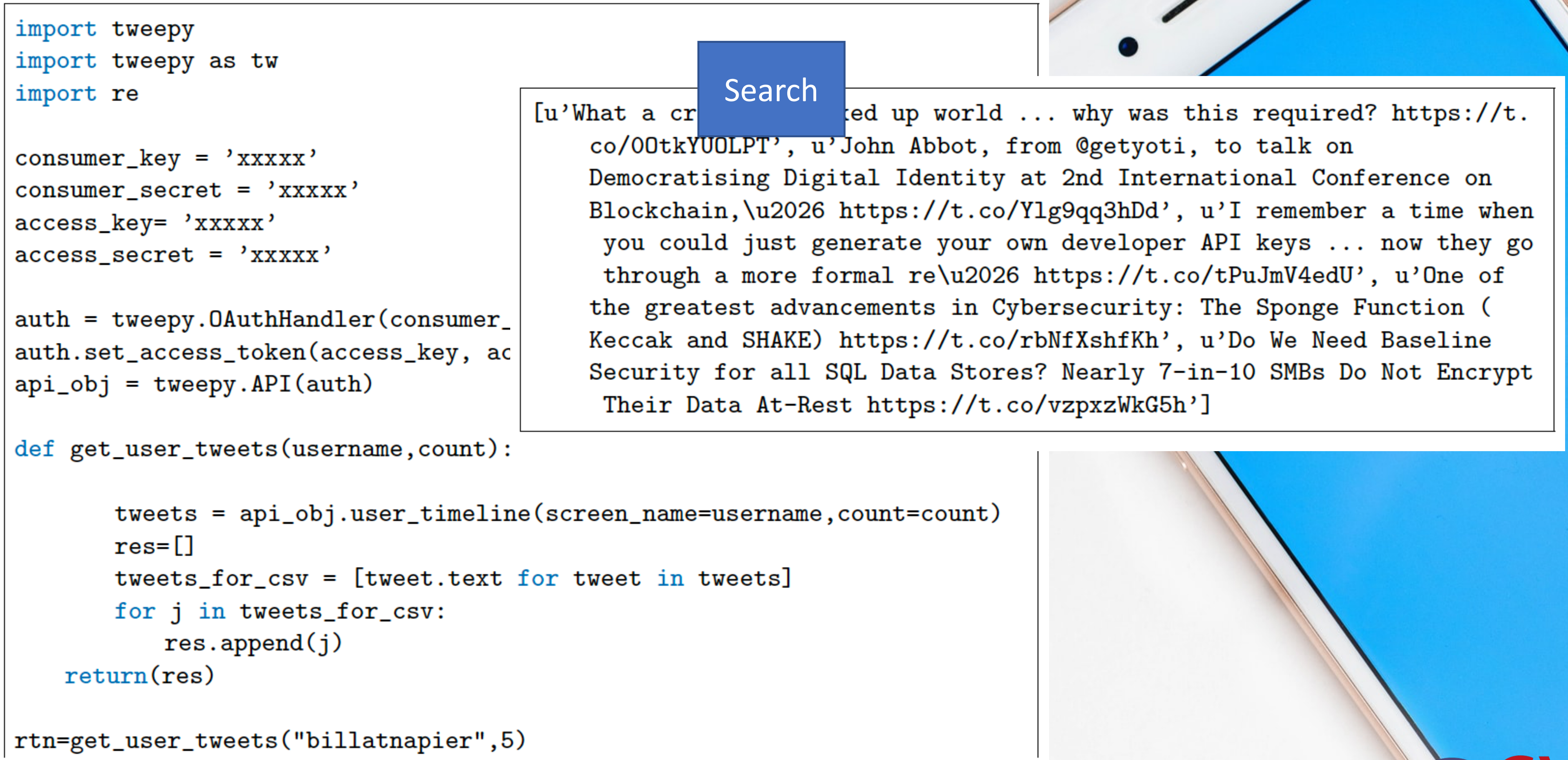
Social Networks





cyber & data

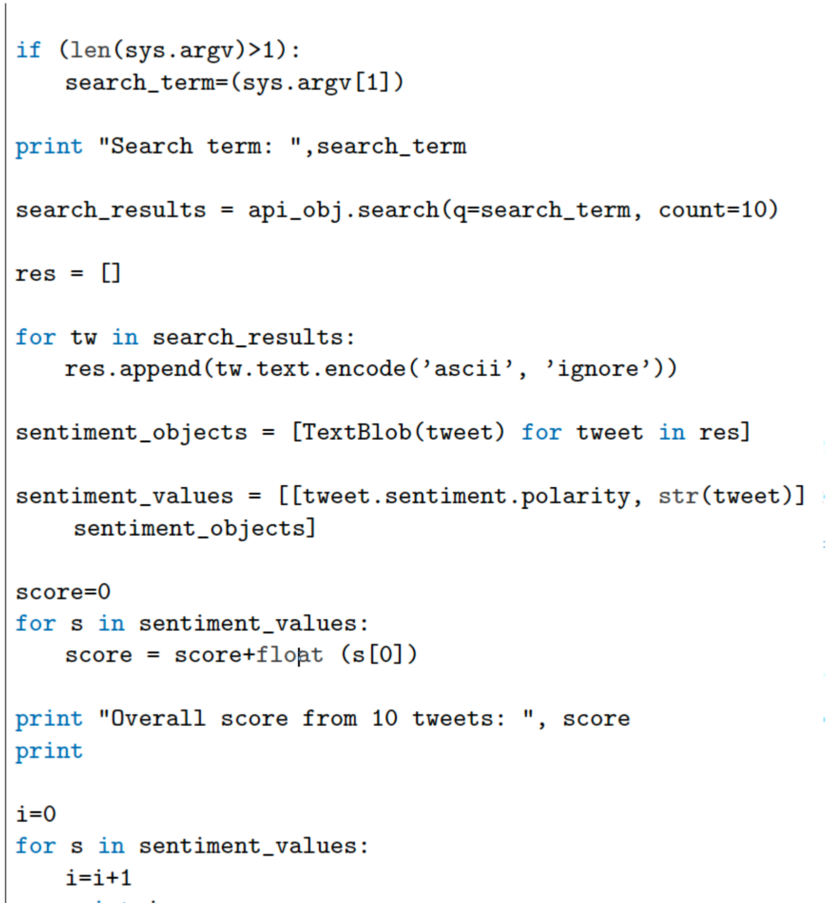
Twitter (User search)



```
import tweepy
import tweepy as tw
import re

consumer_key = ’xxxxx’
consumer_secret = ’xxxxx’
access_key= ’xxxxx’
access_secret = ’xxxxx’

auth = tweepy.OAuthHandler(consumer_
auth.set_access_token(access_key, ac
api_obj = tweepy.API(auth)

def get_user_tweets(username,count):

       tweets = api_obj.user_timeline(screen_name=username,count=count)
       res=[]
       tweets_for_csv = [tweet.text for tweet in tweets]
       for j in tweets_for_csv:
           res.append(j)
    return(res)

rtn=get_user_tweets("billatnapier",5)
```

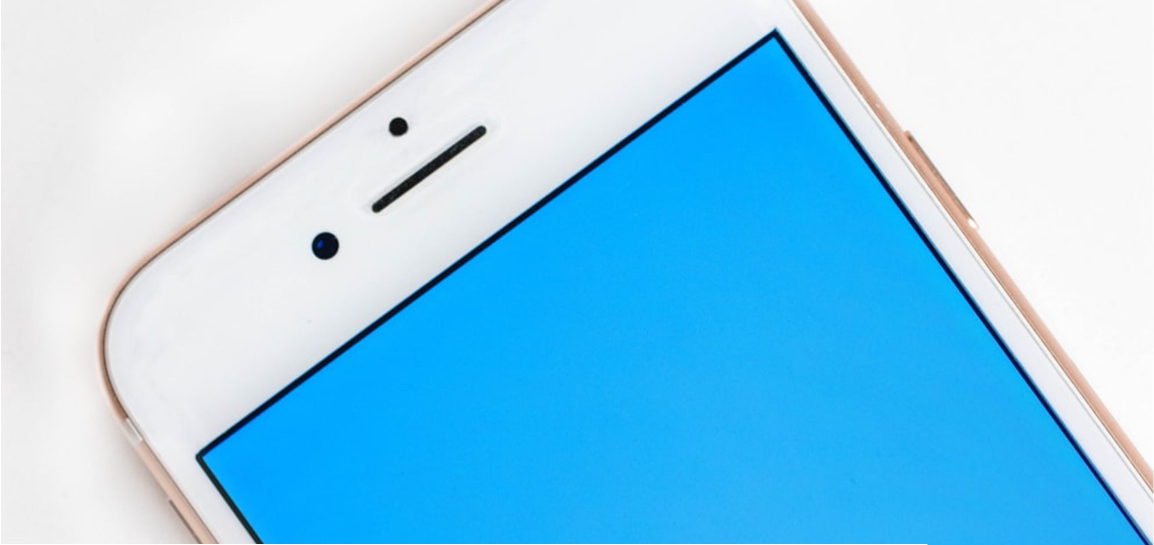
```
[u’What a cr            ked up world ... why was this required? https://t.
    co/OOtkYUOLPT’, u’John Abbot, from @getyoti, to talk on
    Democratising Digital Identity at 2nd International Conference on
    Blockchain,\u2026 https://t.co/Ylg9qq3hDd’, u’I remember a time when
     you could just generate your own developer API keys ... now they go
     through a more formal re\u2026 https://t.co/tPuJmV4edU’, u’One of
    the greatest advancements in Cybersecurity: The Sponge Function (
    Keccak and SHAKE) https://t.co/rbNfXshfKh’, u’Do We Need Baseline
    Security for all SQL Data Stores? Nearly 7-in-10 SMBs Do Not Encrypt
     Their Data At-Rest https://t.co/vzpxzWkG5h’]
```

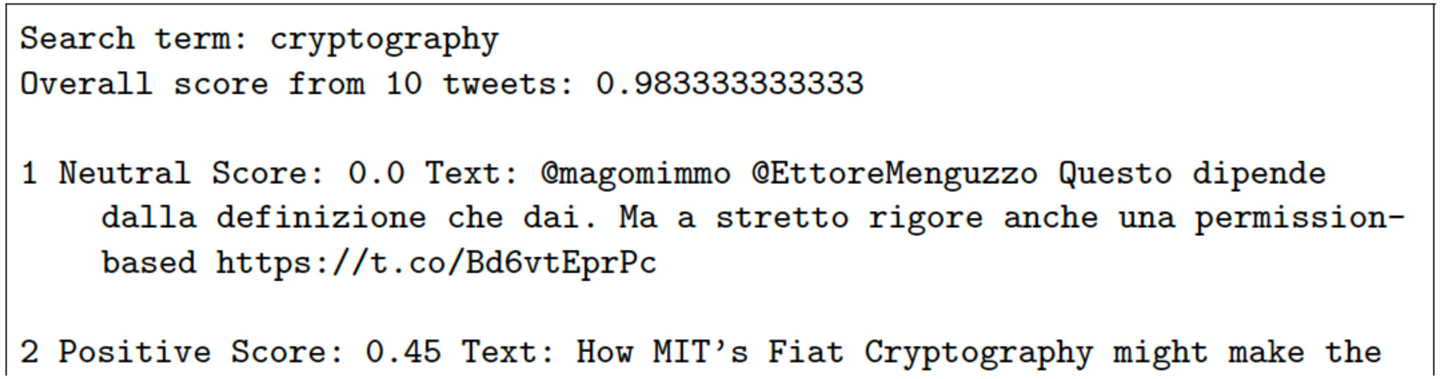
Twitter (Search keyword)



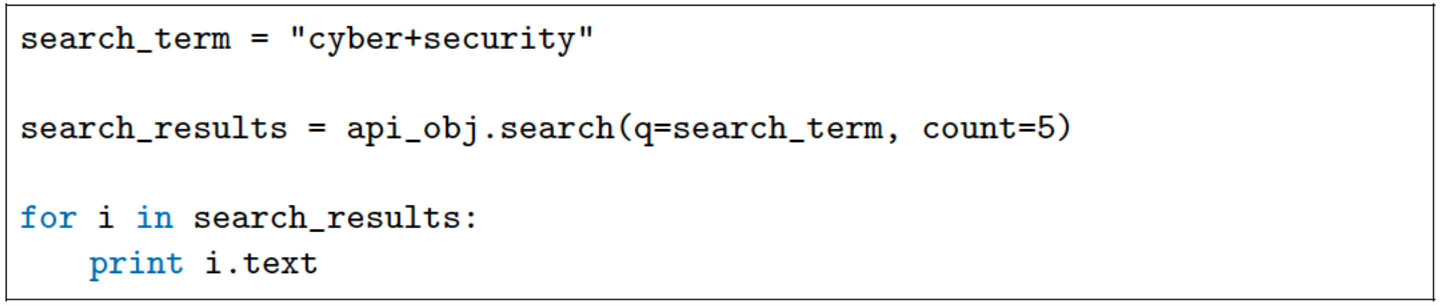
```
if (len(sys.argv)>1):
    search_term=(sys.argv[1])

print "Search term: ",search_term

search_results = api_obj.search(q=search_term, count=10)

res = []

for tw in search_results:
    res.append(tw.text.encode('ascii', 'ignore'))

sentiment_objects = [TextBlob(tweet) for tweet in res]

sentiment_values = [[tweet.sentiment.polarity, str(tweet)]
     sentiment_objects]

score=0
for s in sentiment_values:
    score = score+float (s[0])

print "Overall score from 10 tweets: ", score
print

i=0
for s in sentiment_values:
    i=i+1
    print i,
    if float (s[0]) > 0:
            print 'Positive ',
    elif float(s[0]) == 0:
            print 'Neutral ',
    else:
            print 'Negative ',

    print "Score: ",s[0], "Text: ",s[1]
    print
```

Search



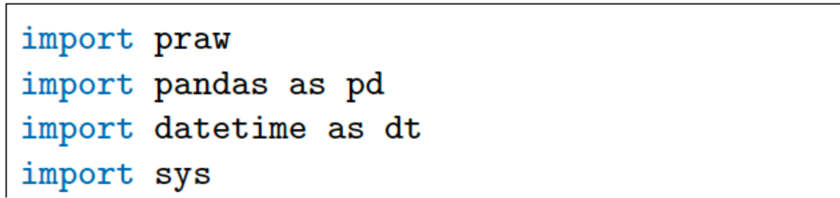


```
Search term: cryptography
Overall score from 10 tweets: 0.983333333333

1 Neutral Score: 0.0 Text: @magomimmo @EttoreMenguzzo Questo dipende
    dalla definizione che dai. Ma a stretto rigore anche una permission-
    based https://t.co/Bd6vtEprPc

2 Positive Score: 0.45 Text: How MIT's Fiat Cryptography might make the
```

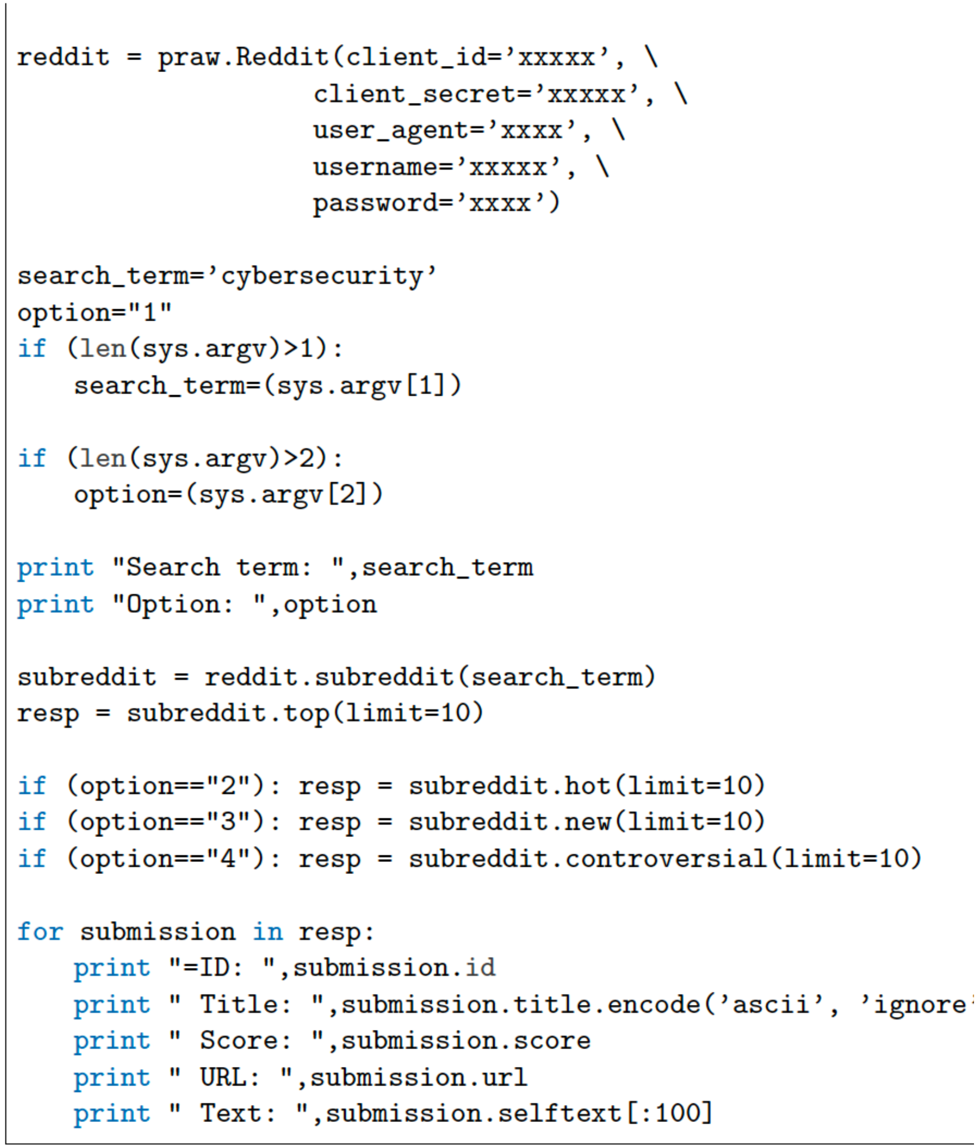


```
search_term = "cyber+security"

search_results = api_obj.search(q=search_term, count=5)

for i in search_results:
    print i.text
```

Reddit

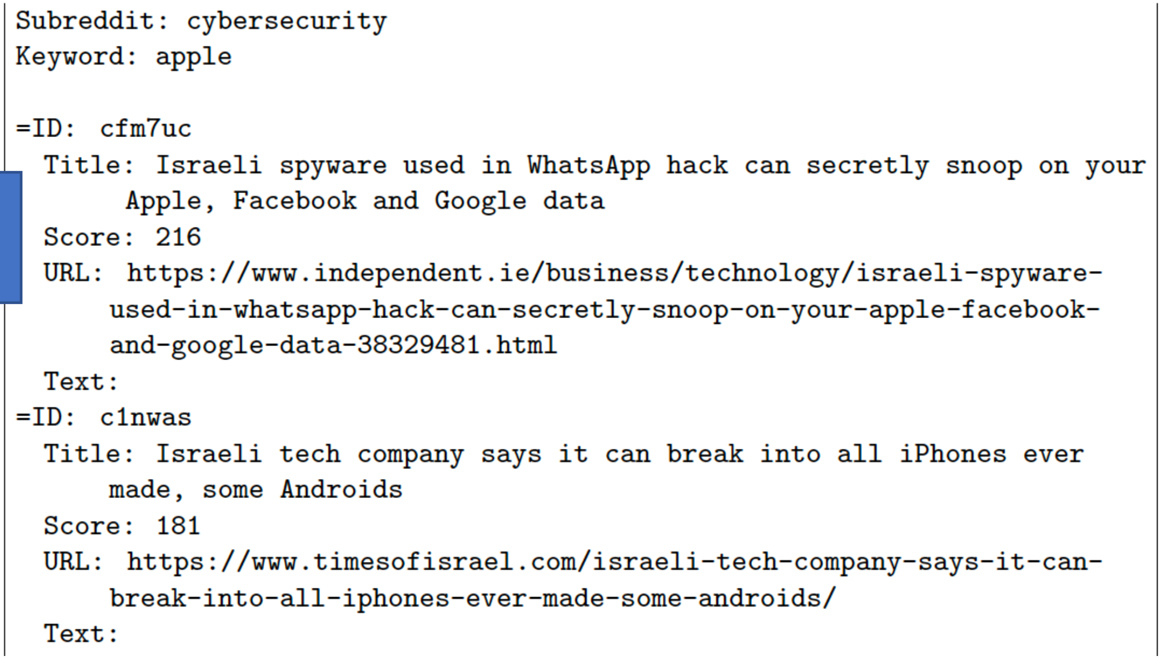


```
import praw
import pandas as pd
import datetime as dt
import sys
```

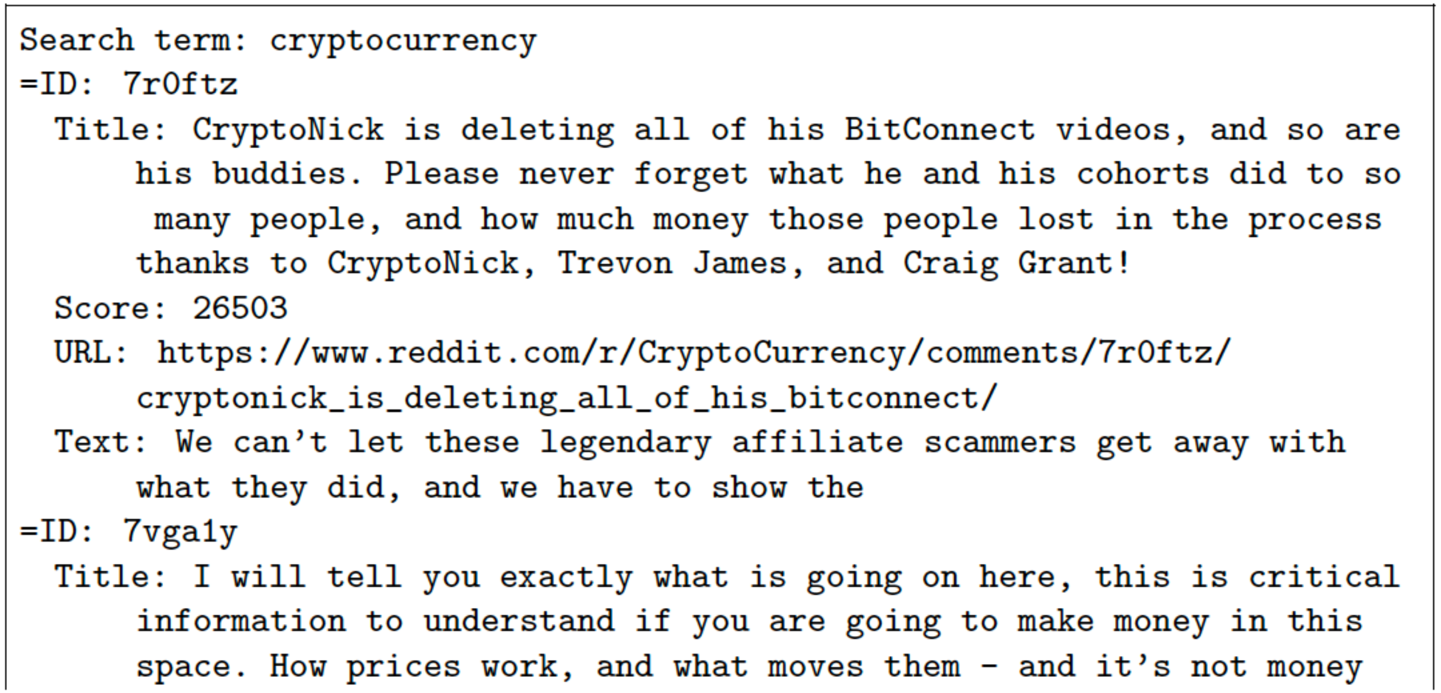


```
reddit = praw.Reddit(client_id='xxxxx', \
                     client_secret='xxxxx', \
                     user_agent='xxxx', \
                     username='xxxxx', \
                     password='xxxx')

search_term='cybersecurity'
option="1"
if (len(sys.argv)>1):
    search_term=(sys.argv[1])

if (len(sys.argv)>2):
    option=(sys.argv[2])

print "Search term: ",search_term
print "Option: ",option

subreddit = reddit.subreddit(search_term)
resp = subreddit.top(limit=10)

if (option=="2"): resp = subreddit.hot(limit=10)
if (option=="3"): resp = subreddit.new(limit=10)
if (option=="4"): resp = subreddit.controversial(limit=10)

for submission in resp:
    print "=ID: ",submission.id
    print " Title: ",submission.title.encode('ascii', 'ignore
    print " Score: ",submission.score
    print " URL: ",submission.url
    print " Text: ",submission.selftext[:100]
```

Search



```
Subreddit: cybersecurity
Keyword: apple

=ID:  cfm7uc
  Title:  Israeli spyware used in WhatsApp hack can secretly snoop on your
        Apple, Facebook and Google data
  Score:  216
  URL:  https://www.independent.ie/business/technology/israeli-spyware-
       used-in-whatsapp-hack-can-secretly-snoop-on-your-apple-facebook-
       and-google-data-38329481.html
  Text:
=ID:  c1nwas
  Title:  Israeli tech company says it can break into all iPhones ever
       made, some Androids
  Score:  181
  URL:  https://www.timesofisrael.com/israeli-tech-company-says-it-can-
       break-into-all-iphones-ever-made-some-androids/
  Text:
```



```
Search term:  cryptocurrency
=ID:  7r0ftz
  Title:  CryptoNick is deleting all of his BitConnect videos, and so are
       his buddies. Please never forget what he and his cohorts did to so
       many people, and how much money those people lost in the process
       thanks to CryptoNick, Trevon James, and Craig Grant!
  Score:  26503
  URL:  https://www.reddit.com/r/CryptoCurrency/comments/7r0ftz/
       cryptonick_is_deleting_all_of_his_bitconnect/
  Text:  We can't let these legendary affiliate scammers get away with
       what they did, and we have to show the
=ID:  7vga1y
  Title:  I will tell you exactly what is going on here, this is critical
       information to understand if you are going to make money in this
       space. How prices work, and what moves them - and it's not money
```

cyber & data

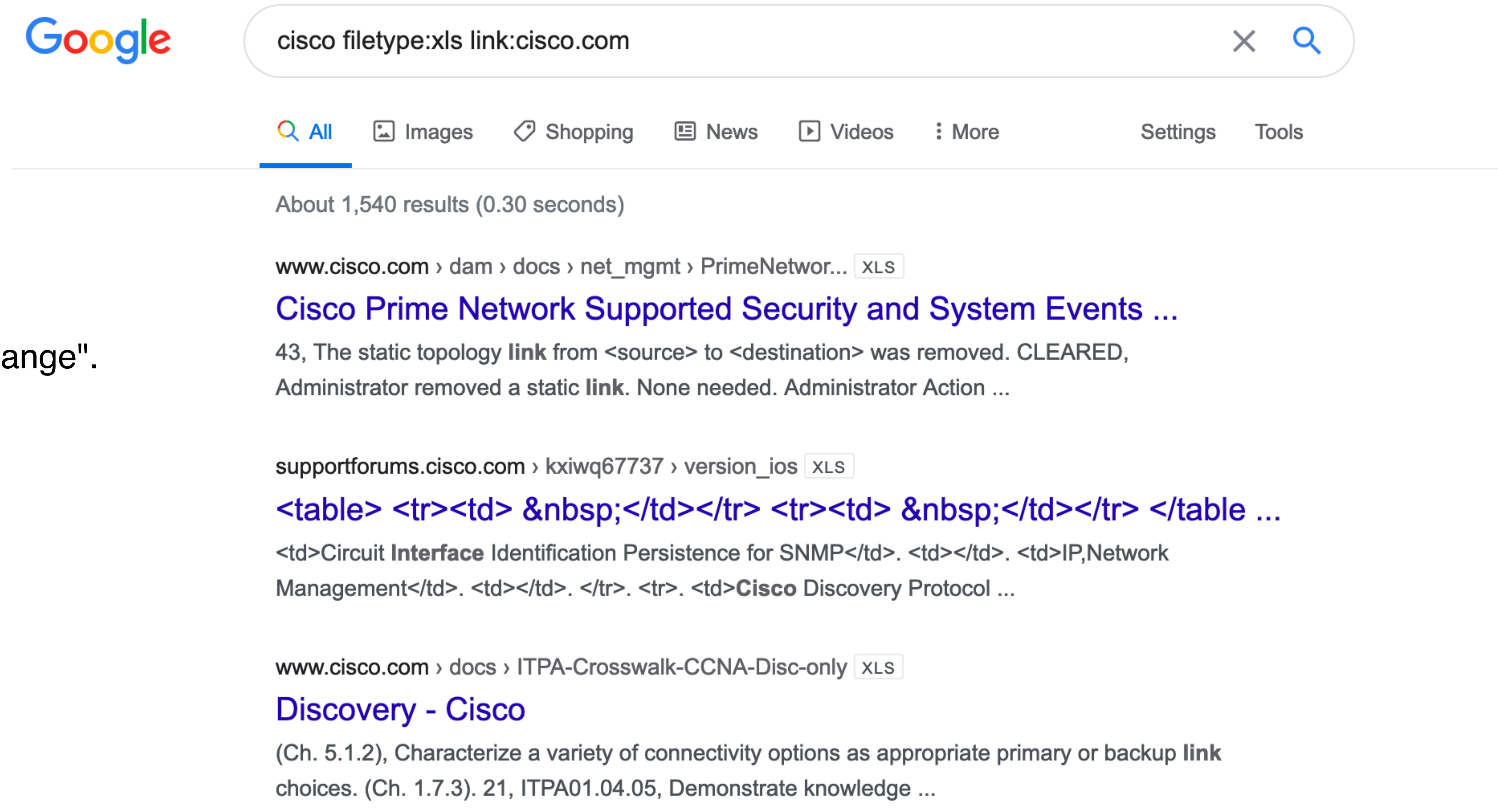
"From bits to information"

Google Hacks

Google Hacks

“fletype:xls
"site:bobco”
“inurl:passwords"
"intext:bob".
"intitle", and "daterange".

Search



cisco filetype:xls link:cisco.com

About 1,540 results (0.30 seconds)

www.cisco.com › dam › docs › net_mgmt › PrimeNetwor... XLS

Cisco Prime Network Supported Security and System Events ...

43, The static topology **link** from <source> to <destination> was removed. CLEARED, Administrator removed a static **link**. None needed. Administrator Action ...

supportforums.cisco.com › kxiwq67737 › version_ios XLS

<table> <tr><td> &nbsp;</td></tr> <tr><td> &nbsp;</td></tr> </table ...

<td>Circuit **Interface** Identification Persistence for SNMP</td>. <td></td>. <td>IP,Network Management</td>. <td></td>. </tr>. <tr>. <td>**Cisco** Discovery Protocol ...

www.cisco.com › docs › ITPA-Crosswalk-CCNA-Disc-only XLS

Discovery - Cisco

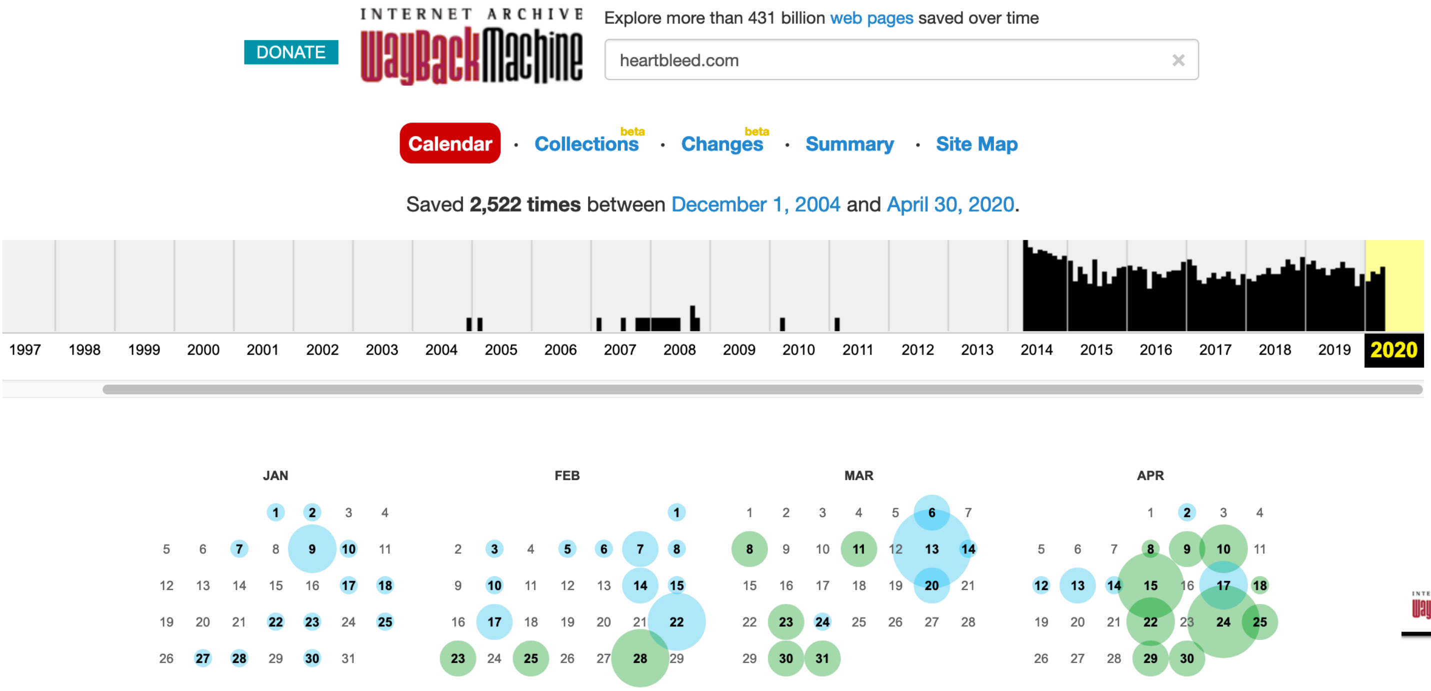
(Ch. 5.1.2), Characterize a variety of connectivity options as appropriate primary or backup **link** choices. (Ch. 1.7.3). 21, ITPA01.04.05, Demonstrate knowledge ...

cyber & data

“From bits to information”

Wayback

Wayback



DONATE

INTERNET ARCHIVE WaybackMachine

Explore more than 431 billion web pages saved over time

heartbleed.com

Calendar · Collections beta · Changes beta · Summary · Site Map

Saved **2,522 times** between December 1, 2004 and April 30, 2020.

JAN

			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

FEB

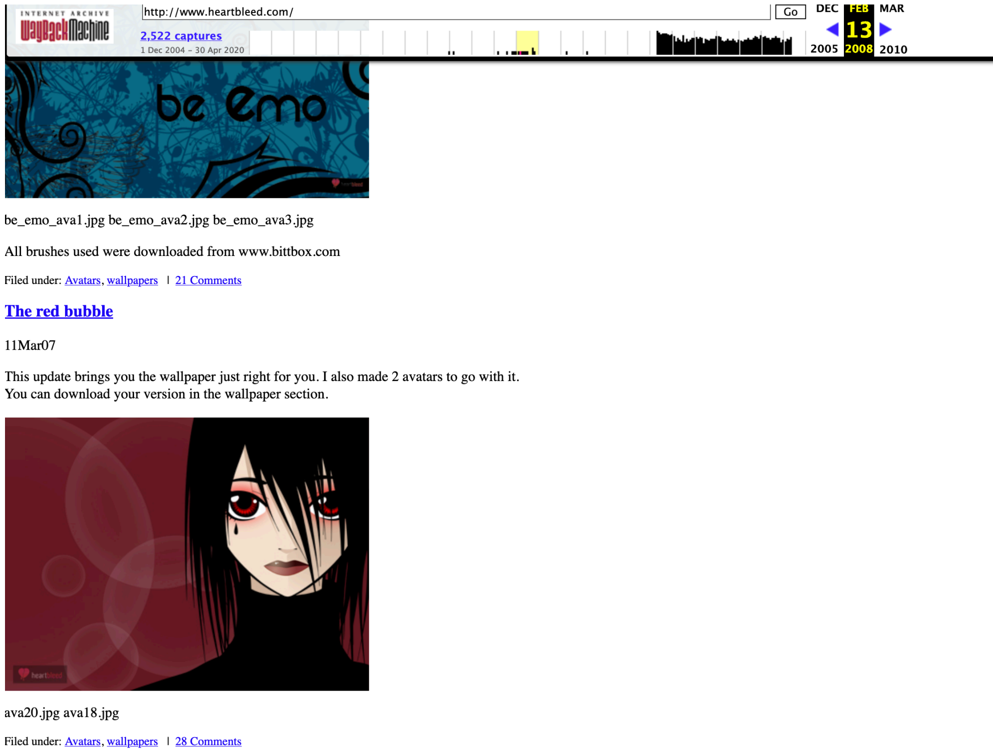
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29

MAR

1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

APR

			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30		



be_emo_ava1.jpg be_emo_ava2.jpg be_emo_ava3.jpg

All brushes used were downloaded from www.bittbox.com

Filed under: Avatars, wallpapers | 21 Comments

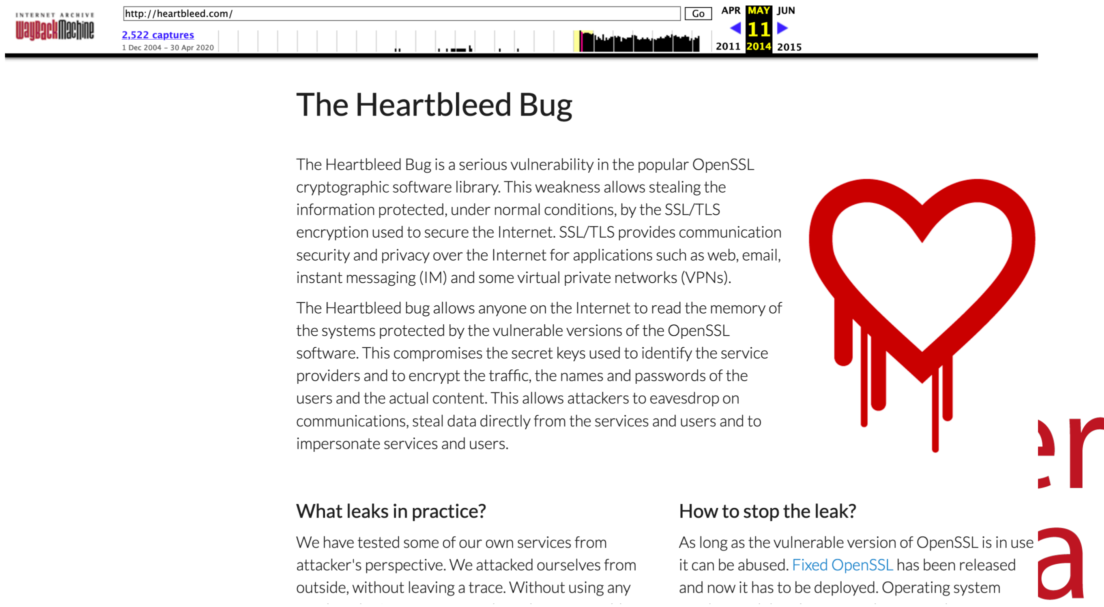
The red bubble

11Mar07

This update brings you the wallpaper just right for you. I also made 2 avatars to go with it. You can download your version in the wallpaper section.

ava20.jpg ava18.jpg

Filed under: Avatars, wallpapers | 28 Comments



The Heartbleed Bug

The Heartbleed Bug is a serious vulnerability in the popular OpenSSL cryptographic software library. This weakness allows stealing the information protected, under normal conditions, by the SSL/TLS encryption used to secure the Internet. SSL/TLS provides communication security and privacy over the Internet for applications such as web, email, instant messaging (IM) and some virtual private networks (VPNs).

The Heartbleed bug allows anyone on the Internet to read the memory of the systems protected by the vulnerable versions of the OpenSSL software. This compromises the secret keys used to identify the service providers and to encrypt the traffic, the names and passwords of the users and the actual content. This allows attackers to eavesdrop on communications, steal data directly from the services and users and to impersonate services and users.

What leaks in practice?

We have tested some of our own services from attacker's perspective. We attacked ourselves from outside, without leaving a trace. Without using any

How to stop the leak?

As long as the vulnerable version of OpenSSL is in use it can be abused. Fixed OpenSSL has been released and now it has to be deployed. Operating system

Search

cyber
& data

"From bits to information"

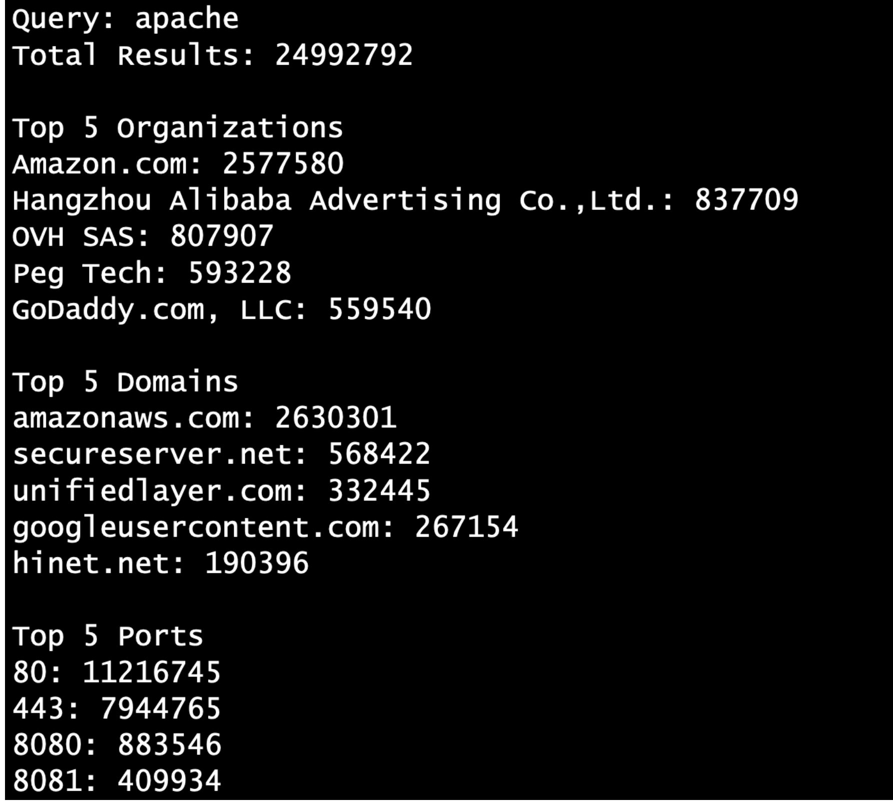
Shodan

Shodan

```
shodan count "port:443"
shodan search "uc-httpd" --limit 10
shodan count "product:apache city:Edinburgh"
shodan search "product:mysql"
shodan search "category:ics country:US"
shodan search "device:webcam city:Edinburgh"
shodan search "port:102"
```

```
shodan search --fields ip_str,port,org "uc-http" --limit 10
shodan search --fields ip_str,org "port:80 os:Linux city:Edinburgh
    country:GB" --limit 10
shodan search --fields ip_str,org "port:21 city:Edinburgh product:nginx"
    --limit 10
```

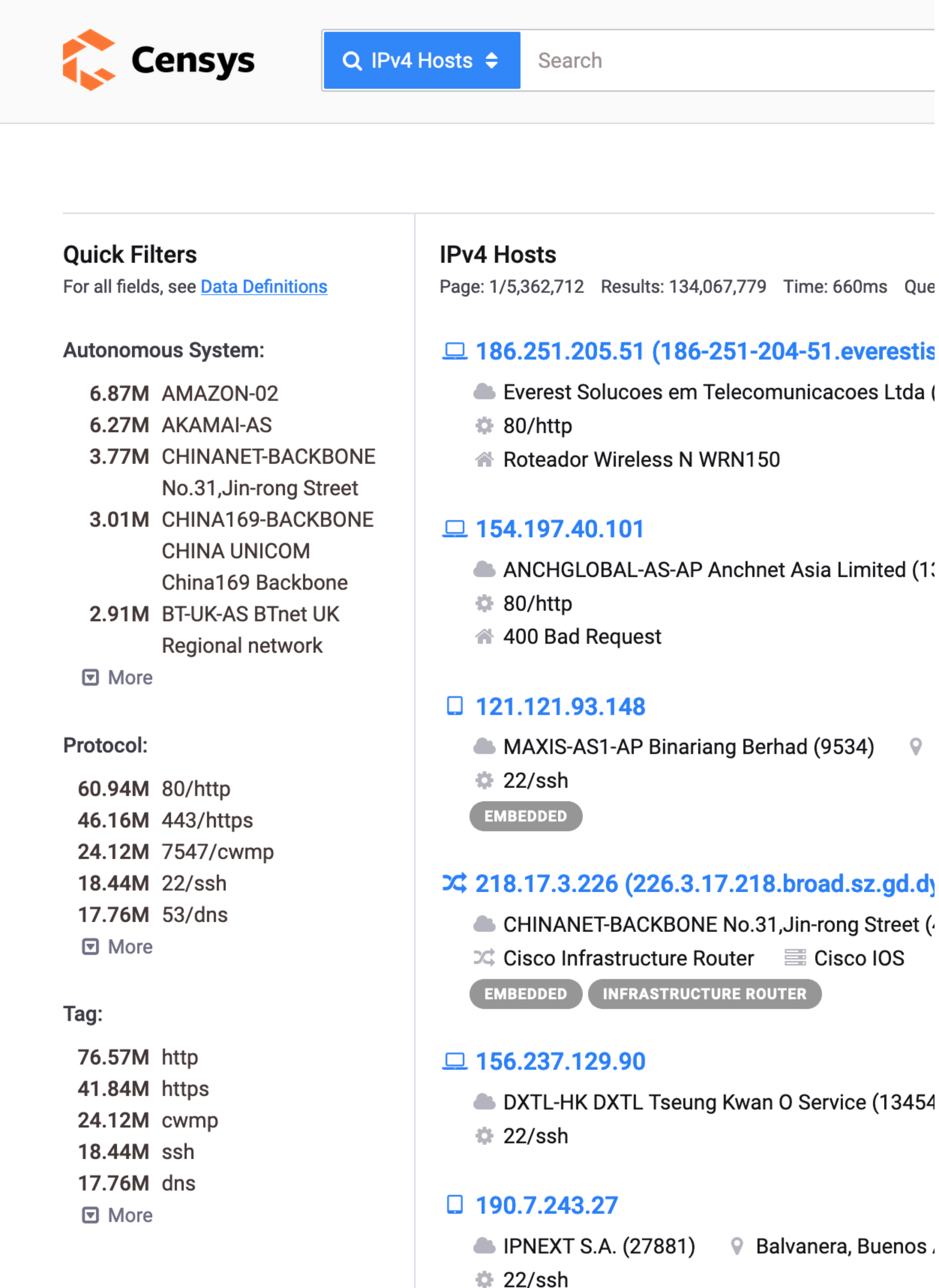
```
shodan stats --facets port device:webcam --limit 5
shodan stats --facets country device:webcam --limit 5
shodan stats --facets port product:apache --limit 5
shodan stats --facets domain product:apache --limit 5
shodan stats --facets org product:apache --limit 5
```



```
Query: apache
Total Results: 24992792

Top 5 Organizations
Amazon.com: 2577580
Hangzhou Alibaba Advertising Co.,Ltd.: 837709
OVH SAS: 807907
Peg Tech: 593228
GoDaddy.com, LLC: 559540

Top 5 Domains
amazonaws.com: 2630301
secureserver.net: 568422
unifiedlayer.com: 332445
googleusercontent.com: 267154
hinet.net: 190396

Top 5 Ports
80: 11216745
443: 7944765
8080: 883546
8081: 409934
```

Search

Censys



Quick Filters

For all fields, see Data Definitions

Autonomous System:

6.87M	AMAZON-02
6.27M	AKAMAI-AS
3.77M	CHINANET-BACKBONE No.31,Jin-rong Street
3.01M	CHINA169-BACKBONE CHINA UNICOM China169 Backbone
2.91M	BT-UK-AS BTnet UK Regional network

More

Protocol:

60.94M	80/http
46.16M	443/https
24.12M	7547/cwmp
18.44M	22/ssh
17.76M	53/dns

More

Tag:

76.57M	http
41.84M	https
24.12M	cwmp
18.44M	ssh
17.76M	dns

More

IPv4 Hosts

Page: 1/5,362,712 Results: 134,067,779 Time: 660ms Que

186.251.205.51 (186-251-204-51.everestis
Everest Solucoes em Telecomunicacoes Ltda (
80/http
Roteador Wireless N WRN150

154.197.40.101
ANCHGLOBAL-AS-AP Anchnet Asia Limited (1:
80/http
400 Bad Request

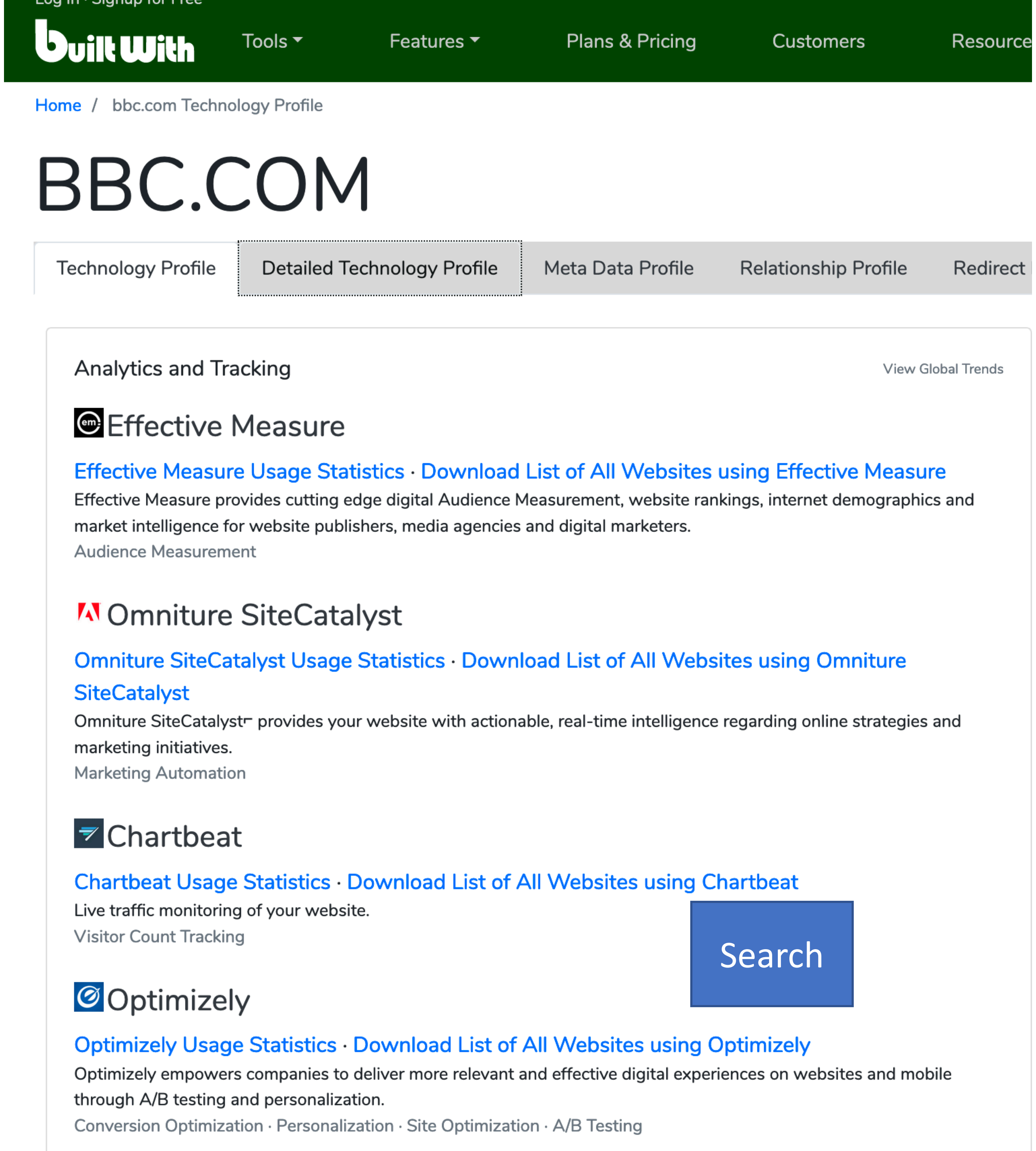
121.121.93.148
MAXIS-AS1-AP Binariang Berhad (9534)
22/ssh
EMBEDDED

218.17.3.226 (226.3.17.218.broad.sz.gd.dy
CHINANET-BACKBONE No.31,Jin-rong Street (
Cisco Infrastructure Router Cisco IOS
EMBEDDED INFRASTRUCTURE ROUTER

156.237.129.90
DXTL-HK DXTL Tseung Kwan O Service (13454
22/ssh

190.7.243.27
IPNEXT S.A. (27881) Balvanera, Buenos
22/ssh

BuiltWith



Log In · Signup for Free

Home / bbc.com Technology Profile

BBC.COM

Analytics and Tracking

View Global Trends

Effective Measure

Effective Measure Usage Statistics · Download List of All Websites using Effective Measure

Effective Measure provides cutting edge digital Audience Measurement, website rankings, internet demographics and market intelligence for website publishers, media agencies and digital marketers.

Audience Measurement

Omniture SiteCatalyst

Omniture SiteCatalyst Usage Statistics · Download List of All Websites using Omniture SiteCatalyst

Omniture SiteCatalyst⌐ provides your website with actionable, real-time intelligence regarding online strategies and marketing initiatives.

Marketing Automation

Chartbeat

Chartbeat Usage Statistics · Download List of All Websites using Chartbeat

Live traffic monitoring of your website.

Visitor Count Tracking

Optimizely

Optimizely Usage Statistics · Download List of All Websites using Optimizely

Optimizely empowers companies to deliver more relevant and effective digital experiences on websites and mobile through A/B testing and personalization.

Conversion Optimization · Personalization · Site Optimization · A/B Testing

Search

&cyber data

&cyber data

"From bits to information"

Open Source Intelligence