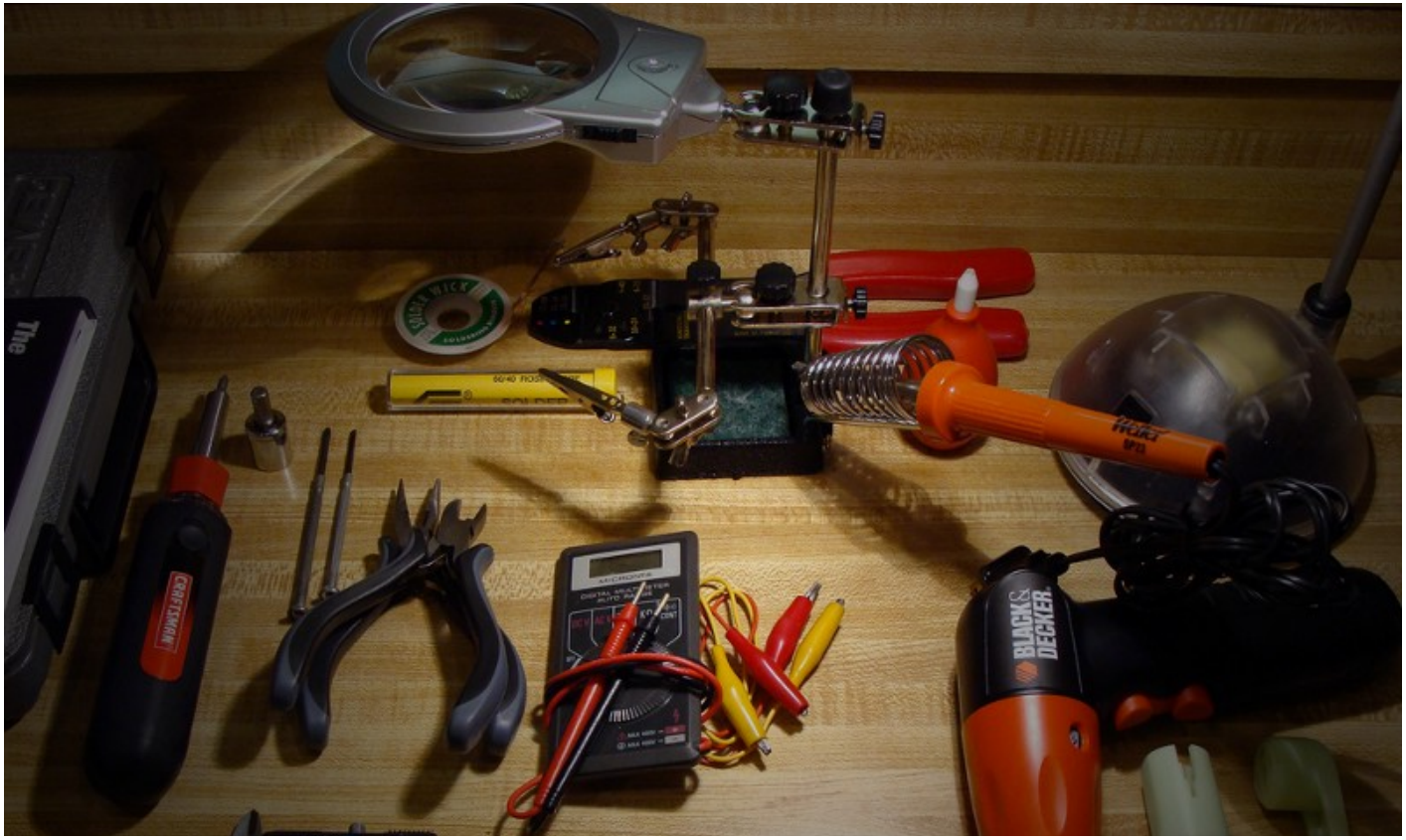
The OSINT Toolkit

Craig Hays Follow
Jul 30, 2016 · 12 min read

Last updated: 11th January 2017



Baidu Maps

http://map.baidu.com/

A Google Maps alternative from the Chinese search company Baidu. Check out the Baidu view of the world in relation to border disputes.

Bing Maps

https://www.bing.com/maps/

Bing Maps are an alternative to the highly popular Google Maps. Bing maps contain details of when its satalite image data was last updated allowing you to judge how accurate the information is.

Bug Bounty Toolkit

https://medium.com/bugbountyhunting/bug-bounty-toolkit-aa36f4365f3f

A collection of tools for network and web application reconnaissance (and attacks).

Camera Trace
http://www.cameratrace.com/

Find additional images taken by the same physical camera using the serial number stored in image EXIF metadata.

Common Crawl
http://commoncrawl.org/ "We build and maintain an open repository of web crawl data that can be accessed and analyzed by anyone. The Common Crawl corpus contains petabytes of data collected over the last 7 years. It contains raw web page data, extracted metadata and text extractions."

Use the Common Crawl API to search their indexed crawl data for sites of interest, then pull the full dataset from their Amazon S3 repository for further analysis. Common Crawl contains snapshots of websites taken over the past 7 years with metadata about the website and services providing it.

Corona

http://corona.cast.uark.edu/

"CORONA is the codename for the United States' first photographic spy satellite mission, in operation from 1960–1972. During that time, CORONA satellites took high-resolution images of most of the earth's surface, with particular emphasis on Soviet bloc countries and other political hotspots in order to monitor military sites and produce maps for the Department of Defense. The more than 800,000 images collected by the CORONA missions remained classified until 1995 when an executive order by President Bill Clinton made them publicly available through the US Geological Survey. Because CORONA images preserve a high-resolution picture of the world as it existed in the 1960s, they constitute a unique resource for researchers and scientists studying environmental change, agriculture, geomorphology, archaeology and other fields."

Cree.py

http://www.geocreepy.com "A Geolocation OSINT Tool. Offers geolocation information gathering through social networking platforms."

Use Cree.py to find the dates, times, and locations of social media posts for a given person from platforms such as Twitter, Instagram, and Flicker. This will allow you to track the movements of named individuals based on their geo-tagged social media posts. Frequent posts from specific locations indicate their own and family/friend's residences.

Dataspoit
https://github.com/upgoingstar/dataspoit

"Performs OSINT on a domain / email / username / phone and find out information from different sources. Correlates and collaborate the results, show them in a consolidated manner. Tries to find out credentials, api-keys, tokens, subdomains, domain history, legacy portals, etc. related to the target. Use specific script / launch automated OSINT for consolidated data. Available in both GUI and Console."

Daum Maps
http://map.daum.net/

A Korean alternative to Google Maps

Folium

https://github.com/python-visualization/folium "Folium builds on the data wrangling strengths of the Python ecosystem and the mapping strengths of the Leaflet.js library. Manipulate your data in Python, then visualize it in on a Leaflet map via Folium."

Use Folium to collate and display location linked data on Leaflet maps through the Python programming language.

FotoForensics

http://fotoforensics.com/

An online EXIF metadata viewer. In addition to the regular metadata it provices an Error Level Analysis (ELA) allowing you to see the compression rate of the image changes highlighting any resaves, crops, stretches, additions, and other modifications.

Geopy

https://github.com/geopy/geopy "A Python 2 and 3 client for several popular geocoding web services. Geopy makes it easy for Python developers to locate

the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources.

Geopy includes geocoder classes for the OpenStreetMap Nominatim, ESRI ArcGIS, Google Geocoding API (V3), Baidu Maps, Bing Maps API, Mapzen Search, Yandex, IGN France, GeoNames, NaviData, OpenMapQuest, What3Words,OpenCage, SmartyStreets, geocoder.us, and GeocodeFarm geocoder services."

Gephy

https://gephi.org/ "Gephi is the leading visualization and exploration software for all kinds of graphs and networks."

Use Gephy to visualise networks and relationships between people, websites, objects, companies, locations, or anything else that is linked. Find first, second, third, etc. degrees of separation between targets.

Gitrob

https://github.com/michenriksen/gitrob

"Gitrob is a command line tool which can help organizations and security professionals find sensitive information lingering in publicly available files on GitHub. The tool will iterate over all public organization and member repositories and match filenames against a range of patterns for files that typically contain sensitive or dangerous information."

Google Chrome

https://www.google.co.uk/chrome/browser/desktop/

Use Google Chrome's developer tools to inspect the source code of websites, view network activity, a very powerful Javascript console, and a variety of plugins to expand it's capability. Use with Hunch.ly and Website IP to cache all content locally along with the IP address of the server displaying the content.

Google Earth

https://earth.google.com/ "lets you fly anywhere on Earth to view satellite imagery, maps, terrain, 3D buildings, from galaxies in outer space to the canyons of the ocean."

Use Google Earth for location reconnaissance. Import your own KML data to plot information against physical locations and view geo-linked data relevant to the location

Google Maps and Google Streetview
https://maps.google.com/ and https://www.google.com/maps/streetview/

Online map tools and 360 images taken from millions of locations around the globe. Use these to gather information about a location without needing to physically be there. Combine with Panoramio, Instagram, Flickr, and Twitter data to build a visual picture of a location from multiple sources and individuals. Useful for finding the locations of images and videos without geo-tag data embedded.

Google URL Shortener
https://goo.gl

Not much use for OSINT alone, but add **.info** to the end of any Google shortened url to view analytics gathered on the link such as number of clicks, demographics, and locations of the visitors.

HERE
https://maps.here.com/

Nokia's alternative to Google Maps.

Hunch.ly
https://www.hunch.ly/ "Inspector Hunchly toils in the background of your web browser to track, analyze and store web pages while you perform online investigations. Forgets nothing, keeps everything. Inspector Hunchly Hates SPAM. You will never get any from us. Ever."

Use Hunch.ly to download, index, and search anything you've ever looked at online. Ever. All content works locally without relying on internet resources so even if it gets deleted you can still see it.

Imagga
http://imagga.com/ "Imagga is an Image Recognition Platform-as-a-Service providing Image Tagging APIs for developers & businesses to build scalable, image intensive cloud apps."

Use Imagga to programatically tag images with descriptors of what is in the image. Create a searchable index for lots of images and identify images which only contain items of interest.

Internet Archive Wayback Machine
https://archive.org/web/ “498 billion web pages saved over time.”

View cached versions of websites over time. Useful for finding information that has since been updated or deleted.

IPinfo.io
https://ipinfo.io/AS36040

Lookup ASNs to find associated IP ranges. Change the AS36040 part of the link to other valid ASNs to see all IP ranges allocated. (Link here is to Google Youtube’s ASN as an example)

Jeffrey’s Image Metadata Viewer
http://regex.info/exif.cgi

An online EXIF metadata viewer with address approximation, colour histograms, and approximate geolocation information.

Leaflet

http://leafletjs.com/ “Leaflet is the leading open-source JavaScript library for mobile-friendly interactive maps.”

Use Leaflet to create interactive web based GIS visualisations using data mined through other tools. Use with Folium.

Magic Cookie IP Range Generator

http://magic-cookie.co.uk/iplist.html

Use this tool to generate a list of IP addresses from an IP CIDR network range such as 192.168.0.0/16. The output can be used as an input for network scanners/scripts that do not take CIDR ranges as an input.

Maltego

https://www.paterva.com/

"Maltego is an extremely powerful OSINT framework, covering infrastructural reconnaissance and personal reconnaissance. The infrastructural component of Maltego enables the gathering of sensitive data about the target organization, email addresses of employees, confidential files which are handled carelessly, internal phone numbers, DNS records, IP address information, geo location of the network, MX servers, and so on. The gathering of such data—known in Maltego as transformations—needs to be creatively and thoughtfully engineered in order to get the best results. Maltego's personal reconnaissance on the other hand helps in the harvesting of person-specific information, such as social networking activity, email addresses, websites associated with the person, telephone numbers, and so on. This happens with the use of search engines on the Internet, which Maltego effectively communicates with to gather all this information."

Maxmind Geolite Database

http://dev.maxmind.com/geoip/geoip2/geolite2/ "GeoLite2 databases are free IP geolocation databases comparable to, but less accurate than, MaxMind's GeoIP2 databases. GeoLite2 databases are updated on the first Tuesday of each month."

Use the Maxmind Geolite databases to map IP addresses to physical locations.

Meanpath

https://meanpath.com/ "A search engine that captures the various bits of code, CSS and HTML across hundreds of millions of websites. This enables you to search for bits of code that might be not be indexed by other search providers."

Use Meanpath to find multiple websites created by the same individual. A Google Analytics account tracking multiple sites will have a pattern such as UA-12345678–10, UA-12345678–11, UA-12345678–12. If custom code has been written by an individual you can search for any other places the individual has used it, or how has copied it from them.

Navar Maps

http://map.naver.com/

A Korean alternative to Google Maps.

NerdyData

https://nerdydata.com/search

Use NerdyData to search for other sites that have included the same source code in their pages. Useful for finding other sites with the same Google Analytics, Adsense, and Affiliate accounts linked to them.

NetworkX

https://networkx.github.io/ "NetworkX is a Python language software package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks."

Use the NetworkX python library to analyse information to identify complex relationships and networks between entities. Export data to GEXF files for rendering and exploration through Gephy.

OnionScan

https://github.com/s-rah/onionscan "The purpose of this tool is to make you a better onion service provider. You owe it to yourself and your users to ensure that attackers cannot easily exploit and deanonymize."

Use OnionScan to find relationships between hidding TOR websites based on hyperlinks, SSH keys, apache headers and status pages, EXIF image metadata tags, software stacks, directory structures, other software versions, and more.

OpenCorporates

https://opencorporates.com/ "The largest open database of companies in the world"

Use the OpenCorporates API to search for public information about companies registered anywhere in the world. View the names of directors and their histories, companies they're linked to, registered company offices, financial statements, and more. Combine with relationship mapping tools (Gephy, NetworkX, etc.) to identify relationships between company directors and companies.

OpenStreetMap

http://www.openstreetmap.org/ "OpenStreetMap is a map of the world, created by people like you and free to use under an open license."

An open source version of Google Maps, edited and updated by anyone.

Panoramio

http://www.panoramio.com/api/data/api.html "Using Panoramio API you can display the photos from Panoramio on your own web site. Geolocated

photos from Panoramio are great to enrich your maps or illustrate information where location is a important factor”

Google’s Panoramio is a geolocation-oriented photo sharing service which adds new images at the end of every month. Use Panoramio to find photographs of specific locations through their published API service or online search. Possible use cases include finding evidence to confirm the locations of scenes in videos by identifying multiple visible landmarks present in videos and in photographs bound to locations. Combine with Google Maps, Google Streetview, and Instagram to further enrich results and strengthen a case.

Pillow

https://pypi.python.org/pypi/Pillow/ “The Python Imaging Library adds image processing capabilities to your Python interpreter. This library provides extensive file format support, an efficient internal representation, and fairly powerful image processing capabilities.”

Use Pillow to manage, manipulate, and process images in bulk using the Python programming language. Combine with other image APIs such as

TinEye and Imagga to quickly tag and search for large numbers of image documents.

Photo Interpretation Student Handbook
http://sites.miis.edu/geospatialtools2012/files/2012/07/Photo-Interpretation-Student-Handbook.-Photo-Interpretation-Principles.pdf

A guide to identifying objects in images from the United States' Defence Mapping Agency.

pyPDF
https://pypi.python.org/pypi/pyPdf/ "A Pure-Python library built as a PDF toolkit".

Use pyPDF for interacting, PDF files. Capable of a variety of operations including opening, scraping, editing, splitting, merging, encrypting and decrypting documents, all through the Python programming language.

Requests library for Python
http://docs.python-requests.org/en/master/ "Requests is the only *Non-GMO* HTTP library for Python, safe for human consumption."

Python does not come with an HTTP library by default. Install the Requests library for all interactions with websites and online services that don't already have their own pre-made python library.

SameID

http://sameid.net/

Use SameID to find websites using the same Google Analytics account based on the UA-1234567-XX tag. Also compares AdSense, Amazon, Clickbank, Addthis services.

SASGIS

http://sasgis.ru/sasplaneta/

SASGIS is a russian made tool for "viewing and downloading high-resolution satellite imagery and conventional maps submitted by such services as the Google Earth , the Google the Maps , the Bing the Maps , DigitalGlobe's , " Kosmosnimki ", Yandex , Yahoo! The Maps , VirtualEarth , Gurtam , by OpenStreetMap , eAtlas , the iPhone maps, maps of the General Staff , and others."

Shodan

https://www.shodan.io/ “Shodan is the world’s first search engine for Internet-connected devices.”

Rather than searching for website content, Shodan allows you to search for the back-end services providing the content. A public API is available allowing you to search straight from your programming language of choice.

SpyOnWeb

http://www.spyonweb.com/

Use SpyOnWeb to find multiple sites owned by the same individual/group based on linked tracking codes for analytics and advertisements.

Stolen Camera Finder

http://www.stolencamerafinder.com/

Find additional images taken by the same physical camera using the serial number stored in image EXIF metadata.

Sublist3r

https://github.com/aboul3la/Sublist3r

Find subdomains of a given domain name using a combined search engine and brute force approach.

Tesseract

https://github.com/tesseract-ocr

Use the Tesseract open source OCR (Optical character recognition) to convert text in images to digital text that can be indexed, searched, and manipulated. The Tesseract OCR application is currently developed by Google. Great for finding information hidden in photographs and scanned documents without having to read each document manually.

Text Mechanic Number Generator

http://textmechanic.com/text-tools/numeration-tools/generate-list-numbers/

Quickly generate a sequential list of numbers for injecting into port scanners/scripts where the specification of ranges is not supported, for

example, a full list of TCP ports from 0–65535.

TimelineJS

http://timeline.knightlab.com/ “Easy-to-make, beautiful timelines.”

Use TimelineJS for creating and and publishing web viewable timelines of events. An effective tool for storytelling and delivering evidence following an investigation.

TinEye

https://www.tineye.com/ “Reverse Image Search. 15.7 billion images indexed and growing.”

Use TinEye to find out where an image came from, how it is being used, if modified versions of the image exist, or to find higher resolution versions. It will also find visually similar images which can provide further evidence that your version has been doctored. The TinEye API allows you to programatically perform image searches saving you plenty of time and manual labour.

Topia.termextract

https://pypi.python.org/pypi/topia.termextract/ “Determines important

terms within a given piece of content. It uses linguistic tools such as Parts-Of-Speech (POS) and some simple statistical analysis to determine the terms and their strength."

Use Topia.termextract to analyse text to identify the core concepts being conveyed within it. Strings of text are scored with numerical values indicating their importance and relevance to the overall message of the text.

TwitterAPI
https://apps.twitter.com/

Use the Twitter API to create your own private apps to search and scrape Tweets by location, hashtag, search strings, and people. Combine with Download follow and follower lists as well as public lists (custom timelines) on a per-user basis. Leaflet and Folium to map to physical locations or with Gephy to identify relationships between individuals.

USGS Earth Explorer
http://earthexplorer.usgs.gov/

WebsiteIP

https://chrome.google.com/webstore/detail/website-ip/ghbmhlgniedlklkpimlibbaoomlpacmk "Simply adds the IP of the website you are viewing to the bottom right."

Use the WebsiteIP Chrome extension to quickly view the IP address of the website you're currently on. More useful that you'd have thought before you installed it, and as it is appended to the body of the HTML document it gets captured in Hunch.ly for later analysis.

Who.is

https://who.is/ "WHOIS Search, Domain Name, Website, and IP Tools"

Useful for identifying the current owner of websites and the servers and ISPs hosting services under a given domain name.

Whois.domaintools.com

http://whois.domaintools.com/

Shows current and historic domain registration information. Includes screenshots and highlights links to other domains registered to the same email

address.

Whoisology

https://whoisology.com/ “Whoisology is a searchable domain name reverse whois / ownership database with over one billion individual domain name records that are updated pretty regularly. Reverse whois is used for cyber crime investigation / InfoSec, corporate intelligence, legal research, business development, and for good ol’ fashioned poking around.”

Another tool for finding current and previous owners of domain names and other domains that they have registered.

Wikimapia

http://wikimapia.org/ “A privately owned open-content collaborative mapping project, that utilizes an interactive “clickable” web map with a geographically-referenced wiki system, with the aim to mark and describe all geographical objects in the world.”

Use Wikimapia to find images and other information for specific locations.

Yandex Maps

https://yandex.com/maps/

Another alternative to Google Maps. Some different information available

yEd Graph Editor

https://www.yworks.com/products/yed "yEd Graph Editor: High quality diagrams made easy"

A free tool for Windows, Linux and OS X for creating network, relationship and other diagrams.

Osint | Open Source | Intelligence | Tools | Security


More from Craig Hays

Recruiting strangers to join your technology startup

Craig Hays
Jul 9, 2018 · 8 min read ★

16

Related reads

VulnHub — Kioptrix: Level 3

Mike Bond
Jun 1, 2018 · 14 min read

233

Related reads

[HTB] Jerry—Write Up

Jio
Nov 18, 2018 · 7 min read

227

Responses

Write a response...