Lesson 08 OSINT Group Activity

Alexander:

Web Scaping

Web scraping is the process of extracting data from websites automatically which saves time and effort in data collection. It allows OSINT users to gather a large amount of information quickly while the user is then able to look at the data and determine what should be done with the findings rather than using the time to manually search through public databases. When talking about web scraping, it is important to note that scraping takes place over social media profiles, public databases, and government sites. Web scraping is an automated process that happens repeatedly, so the information is always up to date. This allows those that need to be aware to act quickly on the new information.

There are multiple popular web scraping tools and methods such as: HTML Parsing, APIs, No-Code Automation, Headless Browsers, OCR Tools. HTML Parsing is extracting specific elements from the website’s HTML structure. APIs are structured data that is sometimes publicly found. No-Code Automation allows people with no technical experience to join in. Headless Browsers simulate user actions to scrape dynamic content. OCR tools extract text from images.

Web scraping and OSINT complement each other because OSINT is the process of gathering information and insights from publicly found information and web scraping is just automating that process so that resources can be used on gaining insights not manually searching for the information.

Some of the ways that they complements each other are: Automated data gathering, data structure and filtering, uncovering patterns in large datasets, combating OSINT challenges with automation. Automated data gathering is what has been discussed already. Data structure and filtering is how the information, that usually comes in raw (unparsed), is extracted and turned into another format like JSON so it is readable. Web scraping can be used to find weird behaviors in the patterns of a specific piece of data. It helps combat OSINT challenges because with automation, it is less likely for human error. Gathering information is a tedious, repetitive task. This ensures that it will be correct every time.

John:

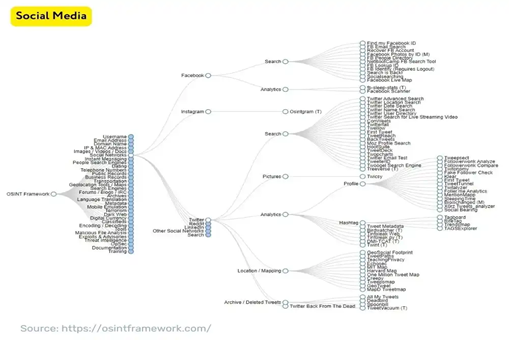
In Open-Source Intelligence (OSINT) social media analysis (SMA) or social network analysis (SNA) gathering is used to collect a variety of information. In this technique, information can be gathered by collecting data, analyzing the data and using the data of publicly available information form social media platforms. Platforms like Facebook, X, LinkedIn, Instagram, YouTube and Reddit hold huge amount of data on personal opinions, behaviors, social networks and real-time events (NEOTAS, n.d.). This technique is a powerful tool for gathering information from public sources on the internet.

Collecting Data is a part of this OSINT process. It involves collecting such information as user profiles, posts and comments, multimedia content, geotagged data and interaction date like shared information and other forms of engagement metrics.

Once the data is collected, that information gets analyzed, through Data Analysis, to pull out the most meaningful information that can be used. All content will be analyzed and examined for the substance of the posts and comments to understand the themes and trends. Through sentiment analysis the information gathered gauges public sentiment and opinion on the various topics.

Finally, through Data Interpretation, is the process of taking the analyzed data and turning it into that can be acted upon immediately. This analysis pulls out Insight Generation, Behavioral Analysis to better understand behavior patterns among users, Predictive Analysis, and then Reporting.

Below is a diagram of what social media analysis may look like:



A use for SNA is through law enforcement. Through the metrics and visual displays of information gathered, police can use SNA to discover, analyze and visualize the social networks of criminal suspects. Using the available data that is gathered, police conduct an examination of an offender’s social network to gain valuable information to the case. Social networks sometimes promote illegal behavior, and they can also provide a source for drug and pornography distribution and international terrorism (Johnson, Reitzel, Norwood, McCoy, Cummings & Tate, 2013)

Works Cited

Johnson, J., Reitzel, J., Norwood, B., McCoy, D., Cummings, D. & Tate, R., (2013, March 5),

Social Network Analysis – A Systematic Approach for Investigating, <https://leb.fbi.gov/articles/featured-articles/social-network-analysis-a-systematic-approach-for-investigating>

NEOTAS, (N.d.), OSINT Sources – Using Social Media for OSINT Investigations,

<https://www.neotas.com/osint-sources-social-media-osint/>

Morgan:

Google Dorking

Google Dorking, also known as google hacking, is an Open-Source Intelligence that uses its own techniques to uncover information throughout the internet. “It leverages the capabilities of Google’s search algorithms to locate specific text strings within search results. (Google Dorking: An introduction for cybersecurity professionals)” Though some worry about it being part of an illegal aspect within the internet, accurately google Dorking is indeed legal. However, when accessing certain files, this might make google Dorking more of a controversy, it “is often utilized by security professionals to identify vulnerabilities in systems . (Google Dorking: An introduction for cybersecurity professionals)”.

Google Dorking works in a way that uses an advanced operation with combining keywords that bring forth a direct search that helps seek specific information. “It can locate files of a particular type, search within a specific website, find keywords in web page titles, or identify pages that link to a particular URL (Google Dorking: An introduction for cybersecurity professionals)”. Google Dorking method finds all public documents that are accessible to anyone, therefore it does not make it illegal when searching the web for such things.

Though there can be dangers when Google Dorking, it is a known terms of service that users should be aware of other legal and ethical boundaries which govern privacy laws. Google Dorking is a valuable tool for revealing data and documents, it is “invaluable tool for practical cyber security research when used responsibly (Cassandra Lee).” Personally, this was a new term for me when it came to researching open-source intelligence. I found it interesting that google of all sources labels a dork as a hacker.

Works Cited

Google Dorking: An introduction for cybersecurity professionals. Splunk. (n.d.). <https://www.splunk.com/en_us/blog/learn/google-dorking.html>

Cassandra Lee, Lee, C., & writer, C. is a. (2024, May 10). Google dorks cheat sheet 2024: How to hack using google. StationX. <https://www.stationx.net/google-dorks-cheat-sheet/>