Alma Mater Studiorum
Università di Bologna
Campus di Cesena

DIPARTIMENTO DI INFORMATICA – SCIENZA E INGEGNERIA
Corso di Laurea in Ingegneria e Scienze Informatiche

OSINT: comparazione e sviluppo di un'integrazione tra i principali strumenti

Elaborato in:
Systems Integration

Relatore:
Vittorio Ghini
Correlatore:
Marco Canducci

Presentata da:
Lucia Fabbri

Sessione III
Anno Accademico 2020-2021

A tutte le persone che amo
e che mi vogliono bene,
anche da lassù ...

Introduzione

Oggigiorno risulta particolarmente semplice per chiunque accedere a grandi quantità di dati ed informazioni in qualsiasi momento utilizzando semplicemente la rete Internet. Con OSINT (Open Source Intelligence) si intende la ricerca e la collezione di informazioni di pubblico dominio, che possono essere reperite nei più svariati modi e sotto varie forme, ed una loro successiva analisi ed elaborazione per trarne specifiche conclusioni.

Tipicamente, le informazioni vengono ricavate interrogando motori di ricerca, ispezionando siti web e social network, oppure informandosi attraverso mezzi di comunicazione come radio, tv, giornali, forum, blog, libri, saggi, video, immagini, interviste e così via.

Data la grande disponibilità di fonti pubbliche reperibili, è importante soffermarci sull'aspetto della coerenza ed attendibilità delle informazioni così come è importante sottolineare che una ricerca completa ed esaustiva non può essere compiuta analizzando ogni singola fonte nei minimi dettagli poiché questo potrebbe comprendere anche informazioni inutili. Devono quindi essere scelte apposite parole chiave, tramite le quali risulti poi essere possibile ricostruire l'essenza del contenuto e stimolare nuove esplorazioni.

Per questi motivi, la ricerca di tutte queste informazioni non viene compiuta a mano dal singolo individuo a cui è affidato il compito, ma avviene con l'ausilio di tecnologie e strumenti specifici scelti a seconda dello scopo della ricerca.

L'obiettivo di questa tesi è quello di sviluppare un programma che possa migliorare ed ampliare la ricerca di informazioni nel web per tutti coloro

che si interfacciano al mondo dell'OSINT. Grazie all'integrazione di singoli strumenti e risorse già note ed ampiamente utilizzate in questo ambito, il programma sviluppato come elaborato di tesi può costituire un punto di riferimento per ogni ricerca poiché permette di scoprire, analizzare, valutare ed utilizzare tutti i dati e le informazioni trovate in un unico momento attraverso un unico strumento.

Al fine di raggiungere questo obiettivo, il progetto possiede specifiche funzionalità, ciascuna delle quali è strettamente legata allo strumento che si decide di utilizzare per effettuare la ricerca. Alcuni strumenti offrono la possibilità di effettuare ricerche interrogando molteplici fonti di informazioni direttamente online, elaborando in autonomia dati di svariata natura e mostrando in output dati organizzati in maniera complessa che devono essere elaborati per una totale comprensione. Attraverso altre tecnologie è possibile focalizzarsi sulla funzionalità di localizzazione e posizionamento così come ci si può concentrare maggiormente sulla ricerca di email, domini, siti web, nome utente utilizzando ulteriori strumenti a disposizione. Ogni informazione che viene trovata può e deve essere analizzata e controllata nella sua validità: in questo contesto viene sviluppata la funzionalità di controllo dei dati. Nel caso in cui si scelga di effettuare la ricerca completa, si potrà invece beneficiare contemporaneamente di tutte le funzionalità appena descritte.

In questa relazione troveremo principalmente tre capitoli, nei quali andremo a descrivere, analizzare e chiarire tutti i vari aspetti e tutte le tematiche legate al progetto di tesi realizzato:

- **Contestualizzazione del problema** è il capitolo con cui si aprirà questo elaborato di tesi, in cui si descriverà il contesto nel quale nasce e si sviluppa la ricerca tramite tecniche di OSINT, quali sono i suoi principali campi di impiego, come questa tipologia di ricerca può essere indirizzata per ottenere determinate informazioni e raggiungere specifici obiettivi, a quali problemi ha dato risposta. Infine, verrà aperta una parentesi circa la sicurezza personale e l'ambito legale nello svolgere e nell'utilizzare questi strumenti di ricerca.

- **Strumenti e tecnologie utilizzate** è il secondo capitolo, nel quale sia andranno a descrivere gli strumenti che sono stati analizzati ed impiegati come fondamenta del progetto, illustrando poi i linguaggi di programmazione, le diverse librerie, le API e tutte le altre tecnologie che sono servite per costruire l'ambiente di sviluppo dell'elaborato.

- **Analisi ed implementazione del progetto** è l'ultimo capitolo in cui si andranno a descrivere in maniera dettagliata il funzionamento dell'intero progetto. Ci si concentrerà poi sulla spiegazione di tutte le sua funzionalità insieme alla descrizione degli strumenti utilizzati per la sua realizzazione, analizzati sia nella complesso dell'elaborato sia ciascuno singolarmente nel suo campo di impiego. Inoltre, per maggiore chiarezza, verranno inserite schermate contenenti il codice sviluppato ed i risultati ottenuti dalla ricerca.

Indice

Elenco delle figure

Capitolo 1

Contestualizzazione del problema

All'interno di questo primo capitolo si andrà a definire che cosa si intende con Open Source Intelligence, quali sono le principali fonti di reperimento dati, in quali campi ed ambienti viene maggiormente impiegata e quali benefici se ne possono trarre; si passerà poi all'analisi della metodologia di ricerca e allo studio degli strumenti e delle tecnologie che vengono utilizzate in queste fasi. Inoltre, si andrà ad analizzare il tutto sotto un punto di vista legale, dalle tecniche di ricerca agli strumenti utilizzati.

Infine, si andrà a studiare lo stato dell'arte attuale circa l'esistenza di progetti simili a quanto sviluppato come elaborato di tesi, sottolineando le eventuali somiglianze o differenze.

1.1 Open Source INTelligence

Per raggiungere i suoi obiettivi e per soddisfare i propri bisogni di qualsiasi natura, l'uomo ha da sempre cercato e tutt'ora cerca di sfruttare e raccogliere tutte le informazioni possibili per utilizzarle a proprio vantaggio. Proprio in questo contesto si può iniziare a parlare di OSINT, acronimo di Open Source INTelligence, ovvero intelligenza delle fonti aperte.

La nascita e l'utilizzo del termine Open Source Intelligence come lo intendiamo al giorno d'oggi è da ricercarsi negli anni che seguirono la guerra fredda, grazie allo sviluppo della rete internet su larga scala e al rapido incremento di informazioni disponibili che si potevano trovare al suo interno. La disciplina moderna dell'Open Source INTelligence si occupa infatti della ricerca e della collezione di informazioni di pubblico dominio tratte da fonti aperte e libere, cioè tutte quelle fonti di dominio pubblico accessibili a chiunque, in maniera gratuita o a pagamento, non sottoposte ad un regime di riservatezza, reperite nei più svariati modi e sotto varie forme. Ma che cosa si intende con pubblico dominio? L'espressione pubblico dominio indica in generale il complesso e la globalità delle opere, in particolar modo delle informazioni che, una volta decorso il termine della protezione legale, possono essere liberamente utilizzate, senza dover chiedere autorizzazioni al titolare delle informazioni o a terzi e senza dover corrispondere un qualche compenso.[19]

Seguendo questa definizione, si può quindi capire perché l'Open Source INTelligence venga anche definita come "Intelligenza *delle* fonti aperte" e non Intelligenza *dalle* fonti aperte, come ci si potrebbe aspettare. Nel primo caso, infatti, si andrebbe ad analizzare tutto ciò che le fonti propongono e mostrano, ovvero le informazioni; nel secondo caso l'argomento di studio ed interesse sono proprio le fonti stesse, le quali devono subire un processo di analisi, pulizia e controllo prima di poter essere considerate attendibili. Infatti, prima di poter essere utilizzate, viene studiata la natura della fonte insieme alle sue caratteristiche, in modo tale da poter costruire e strutturare un archivio di fonti valide, ben organizzate e messe in relazione tra loro, dalle quali poter estratte informazioni corrette e veritiere. [20].

Così dicendo possiamo già intuire che cosa prevede la metodologia di ricerca OSINT. In una prima fase si raccolgono quindi tutte le possibili fonti di informazioni e da esse si estrapolano tutti i dati, anche quelli più nascosti, in modo da poter ricostruire un contenuto in maniera coerente e corretta partendo dai singoli mattoncini d'informazione trovati in precedenza nei canali

più disparati. Successivamente, le informazioni trovate vengono analizzate ed elaborate per trarne specifiche conclusioni a seconda dell'obiettivo da raggiungere. Utilizzando altre parole, si può dire che l'OSINT consente il reperimento di contenuti 'potenzialmente' informativi su fonti aperte a tutti senza l'utilizzo di metodi coercitivi.

1.2 Metodologia di ricerca dell'OSINT

Quando si parla di metodologia di ricerca, si intende tutto l'insieme di tecniche e di procedure che vengono utilizzate ed applicate in maniera sistematica durante le fasi di studio, analisi, comprensione ed elaborazione di un determinato problema, in modo tale da garantire validità e rigore scientifico ai risultati ottenuti. [21] Considerando la grande vastità di fonti, contenuti ed informazioni disponibili, è fondamentale approcciarsi all'OSINT attraverso una specifica metodologia di lavoro. In questo modo, da una serie di dati grezzi, generici e senza collegamenti tra loro si passa a contenuti attentamente analizzati, filtrati e controllati, destinati a soddisfare la richiesta iniziale. In dottrina, per riassumere le fasi della metodologia di ricerca, si parla delle cosiddette "quattro D":

- *Discovery*: individuazione generale di tutte le fonti di possibile interesse per raggiungere l'obiettivo.
- *Discrimination*: selezione e catalogazione delle fonti più utili ed attendibili.
- *Distillation*: ulteriore filtraggio delle sole informazioni rilevanti per la ricerca.
- *Dissemination*: diffusione dei risultati ottenuti dal processo di ricerca OSINT ai soggetti interessati.

In maniera più ampia e dettagliata, la metodologia di ricerca OSINT può essere riassunta utilizzando la figura sottostante, immaginandola anche come

un processo ciclo, da poter rieseguire più e più volte in modo da ottenere risultati sempre più accurati.

Figura 1.1: Metodologia di ricerca dell'OSINT[1]

Si andranno ora ad analizzare brevemente tutte le fasi illustrate nell'immagine sovrastante.

- **Direction and planning**: si definisce il target su cui si andrà a lavorare, lo scopo della ricerca e gli obiettivi finali da raggiungere.
- **Collection**: inizia la fase di ricerca raccolta delle informazioni, tipicamente sul web data l'enorme quantità di dati pubblici disponibili, ma anche tramite social media, indirizzi email o dati personali, nomi a dominio e altro.
- **Processing and collation**: avvalendosi di specifiche tecnologie e specifici strumenti, insieme ad adeguate tecniche di ricerca, tutte le informazioni raccolte vengono elaborate e confrontate tra loro per produrre nuovi contenuti utili al raggiungimento dell'obiettivo.

- **Analysis and integration**: la vera e propria fase di analisi, in cui i dati raccolti non vengono più analizzati singolarmente ma vengono ricercate eventuali analogie, differenze o collegamenti tra di essi. Applicando basilari tecniche di analisi lessicale, geospaziale o semantica ai dati in ingresso relative al target, è possibile estrarre dati che possono essere ricondotti a informazioni personali, di natura organizzativa, relative alla rete e altro. Queste informazioni vengono a loro volta rielaborate grazie all'ausilio di tecnologie come data mining e intelligenza artificiale che facilitano l'impiego di metodi classificazione, regressione, rilevamento di pattern o anomalie.

- **Production and dissemination**: in questa ultima fase di ricerca, si hanno a disposizione una serie di nuove informazioni derivate dall'analisi e dell'elaborazione dei contenuti raccolti in precedenza. Se quanto ottenuto non soddisfa le richieste prestabilite, tutti i passaggi appena elencati possono essere ripetuti ciclicamente per raffinare le ricerche.

Riassumendo, da una prima fase di analisi e definizione degli obiettivi che si vogliono raggiungere, si passa alla fase di raccolta dati. Tutte queste informazioni verranno poi analizzate ed elaborate per poterne estrarre e derivarne conclusioni soddisfacenti. Questo processo può essere reiterato più e più volte fino al raggiungimento degli obiettivi finali prestabiliti durante la prima fase.

1.3 Fonti di reperimento dati ed informazioni

Alla base dell'attività di ricerca dell'OSINT si trovano i dati e le informazioni di pubblico dominio. Seguendo la logica della definizione di OSINT, non è difficile pensare a come le singole fonti di reperimento siano radicalmente cambiate con l'avanzare del tempo e con il progredire della società.

Storicamente, la principale fonte d'informazione era costituita da tutti quei mezzi di comunicazione che oggigiorno consideriamo "tradizionali" come

televisione, radio, quotidiani, discorsi pubblici e conferenze stampa, tramite le quali si poteva avere l'accesso a molteplici dati pubblici (open data). In aggiunta, si potevano ricercare informazioni anche tramite fonti secondarie come pubblicazioni scientifiche, libri, rapporti governativi, dati demografici, database istituzionali. Infine, anche la cosiddetta letteratura grigia si prestava ad essere un'utile fonte di informazione, in quanto si poteva accedere a tutto quel materiale non pubblicamente disponibile ma diffuso solo in ambienti ristretti.

Agli albori della disciplina dell'Open Source Intelligence, le tecnologie, gli strumenti e i mezzi di comunicazione disponibili per effettuare le ricerche erano limitati a queste. Tutto ciò che veniva ritrovato era direttamente comprensibile ed utilizzabile dagli utenti che avrebbero effettuato un'operazione di ricerca poiché non sorgeva la necessità di rielaborare e semplificare il tutto prima del suo utilizzo.

Nel corso del tempo però, grazie alla diffusione e alla possibilità quasi omogenea di poter accedere alle rete Internet, il mondo dell'informazione e della ricerca sono profondamente cambiati. L'evoluzione delle strutture di comunicazione ha portato alla creazione di un nuovo flusso di informazioni derivanti dall'utilizzo della rete e, insieme a queste nuovo flusso, è nata la l'esigenza di sviluppare nuove tecniche nel reperimento dei dati e, di conseguenza, adottare un nuovo approccio verso di esse.

Al giorno d'oggi infatti, con la crescita esplosiva delle comunicazioni web e l'enorme volume di dati digitali prodotti in tutto il mondo, ci si può focalizzare anche sul reperimento delle informazioni su fonti aperte che si trovano online. Analizzando tutte le pubblicazioni che si trovano nel web, nei social network, in post e discussioni su blog o forum, si possono trovare innumerevoli contenuti generati dagli utenti in maniera più o meno consapevole. In aggiunta, risulta molto facile ricercare dati o singoli dettagli a partire da essi oppure avendo a disposizione file multimediali, audio, immagini o video digitali.

Considerando le vaste potenzialità del web e l'insieme di tutti i contenuti

di cui è formato, spesso è richiesta una ricerca più approfondita per andare a trovare tutti quei dati provenienti da fonti nascoste e ai quali risulta più difficile risalire:

- **Dark web e deep web**: qui si trova tutto ciò che non viene indicizzato dai comuni motori di ricerca, a cui si può accedere talvolta con un normale browser ma più frequentemente utilizzando specifici software per consentire una navigazione anonima e proteggere la propria privacy ed identità.

- **Metadati ed indirizzi IP**: forniscono informazioni circa la natura del contenuto e quindi hanno una forte legame con la fonte di informazione di provenienza.

- **Informazioni geospaziali**: per rilevare una specifica posizione si possono utilizzare fotografie e mappe satellitari o altre tecnologie come giroscopi, accelerometri, sonde. Con lo sviluppo dei social network e l'utilizzo di app mobile, la posizione in tempo reale di un utente risulta essere sempre esposta e facilmente individuabile.

Un'altra imprescindibile fonte di informazione è ottenuta sfruttando il servizio di *intelligence umana*. In questo caso, si utilizza lo conoscenza diretta delle persone nello specifico campo e ambiente di interesse unito a tutta la rete di conoscenze o di eventi ad essi collegati, ottenendo quindi informazioni di vario tipo che non si sarebbero potute reperire se non sfruttando queste relazioni.

Focalizzandosi proprio sulla definizione di "Open Source", è possibile effettuare una distinzione riguardante l'origine delle fonti trovate e la natura delle informazioni che da esse si ricavano:

- *Open Source Data (Dati da Fonti Aperte)*: l'insieme dei dati ottenuti da fonti aperte, ovvero il sistema di fonti primarie grezze a cui fare riferimento per iniziare le ricerche.

- *Open Source Informations (Informazioni da Fonti Aperte)*: l'insieme di tutti i dati secondari emersi dal processo di operazioni di analisi, filtraggio e verifica delle Open Source Data.

- *Open Source INTelligence (OSINT)*: l'insieme di tutte le informazioni derivanti da un processo di ricerca, selezione, filtraggio e condivisione con specifici destinatari, in modo da rispondere ai loro bisogni informativi utilizzando gli strumenti propri dell'intelligence.

- *Open Source INTelligence Verificate (OSINT-V)*: l'insieme di tutti i dati, i contenuti e le informazioni trovate che sono state sottoposte a ulteriori controlli di validità, veridicità e quindi risultano essere fonti attendibili.

Osservando la seguente immagine, è possibile riassumere in maniera facile, veloce ed intuitiva le principali fonti di reperimento dati ed informazioni nell'ambito dell'OSINT, descritte precedentemente. Più precisamente, nel cerchio più esterno si trova la tipologia di fonte ricercata mentre nel cerchio più interno vengono riportati esempi concreti da cui poter recuperare quella determinata tipologia di dato.

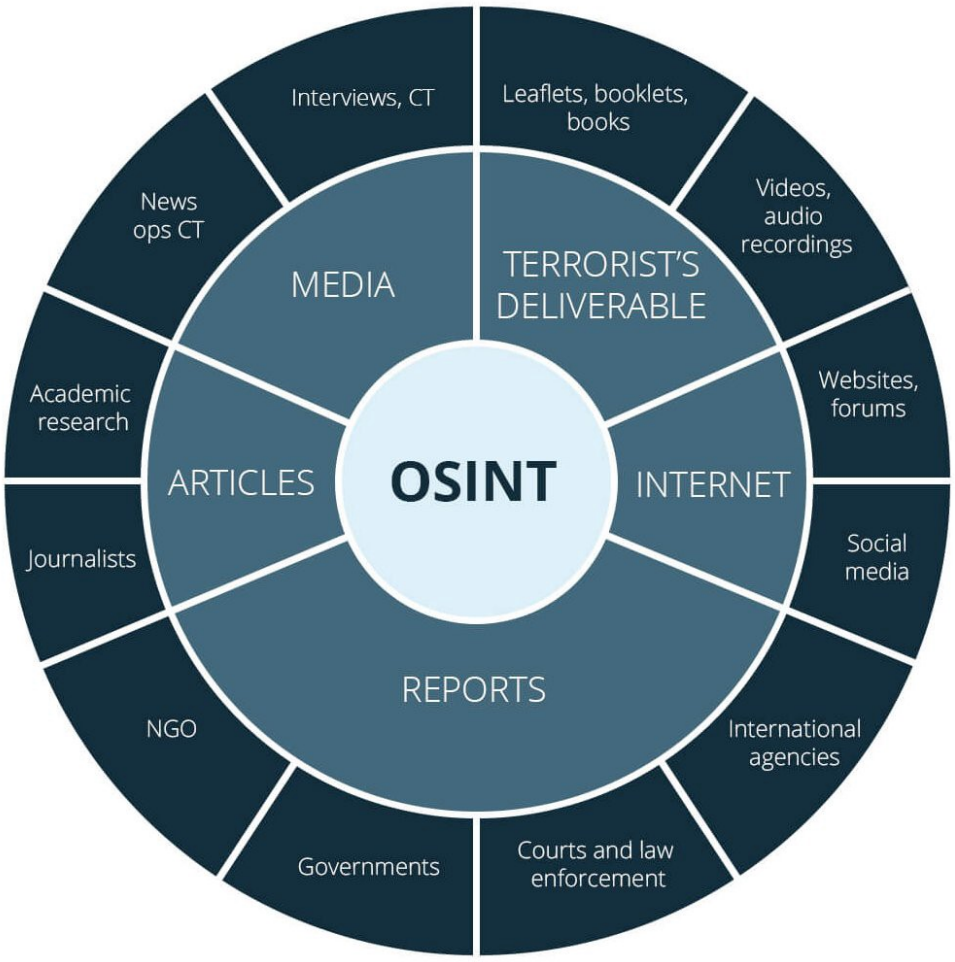


Figura 1.2: Principali fonti di reperimento dati ed informazioni[2]

1.3.1 Attendibilità delle fonti e delle informazioni

A tutto questo grande calderone di dati ed informazioni, non ci si può però avvicinare senza un minimo di sospetto o diffidenza. Infatti, data la grande disponibilità di fonti pubbliche reperibili, è importante soffermarci sull'aspetto della coerenza e attendibilità delle informazioni, poiché non tutto ciò che troviamo durante la fase di ricerca può essere utile in fase di analisi se non rispecchia la realtà dei fatti.

Se si considerano e si analizzano attentamente tutti i contenuti che ci vengono forniti tramite i social media, i blog, i forum oppure altre fonti di

informazione meno autoritarie, non si può fare a meno di pensare a come questi contenuti siano fortemente plasmati e frequentemente influenzati dalle opinioni soggettive di chi le pubblica. A differenza delle fonti chiuse, le quali sono connotate da un maggiore livello coerenza grazie ad una attenta cura e ad una selezione di affidabilità da parte di chi se ne occupa, le fonti di informazione aperte che utilizza l'OSINT non sono di per sé sempre inaffidabili ma è necessario eseguire svariati controlli al fine di garantire la genuinità, l'inalterabilità e l'attendibilità della fonte stessa. Inoltre, se si considera anche la continua evoluzione logaritmica dei contenuti digitali, molti elementi che sono contenuti nel web non vengono neppure indicizzati a causa di una sovra-saturazione di informazione. Gli addetti alle ricerche, si trovano sempre più spesso a dover fare i conti con l'impossibilità di vedere con i propri occhi tutti i contenuti digitali disponibili, così come risulta difficile controllare e verificare autore, data di notizie oppure superare barriere linguistiche.

Qualora questi controlli non venissero eseguiti, il rischio di fare affidamento su contenuti non attendibili o addirittura distorti è molto elevato. Un semplice esempio di ciò, lo si può riscontrare nelle numerose fake news che ogni giorno invadono il mondo del web, camuffate come notizie veritiere ed appetibili, che girano nella rete influenzando e manipolando comportamenti e pensieri altri oppure manipolando emozioni e sentimenti degli utenti che si approcciano ad esse: si parla in questo caso di *Sentiment Analisys* Senza un'attenta valutazione dei contenuti di queste notizie, i risultati che si possono ottenere da una ricerca di questo tipo risulteranno essere sicuramente compromessi o quantomeno influenzati negativamente.

È quindi bene osservare che le informazioni possono essere reperite sia da fonti autorevoli che da fonti meno autorevoli, ma in quest'ultimo caso bisogna prestare particolare attenzione a tutto ciò che si intende utilizzare: attraverso l'utilizzo di determinate linee guida, rispettando le best practice per la gestione e la raccolta di prove e seguendo determinate tecniche di ricerca è possibile superare lo scoglio dell'attendibilità e della veridicità delle informazioni raccolte.

A questo proposito, non si può fare a meno di evidenziare come le fonti, le informazioni, i dati e tutti i contenuti ritrovati durante una ricerca debbano sottostare a rigidi protocolli che utilizzano metodologie e tecniche di valutazione prima di essere utilizzati in un contesto istituzionale, giudiziario e di ricerca. In materia di fonti aperte, la comunità scientifica ha imposto diversi protocolli per la validazione dell'attendibilità di fonti ed informazioni come mostrato nelle seguenti immagini: la prima immagine rappresenta i codici di attendibilità di una fonte ed i codici di attendibilità della valutazione di un'informazione, rispettivamente in una scala da A a D e da 1 a 4, come stabilito dall'EUROPOL; la seconda immagine rappresenta invece l'affidabilità di una fonte e di un'informazione secondo la NATO, rispettivamente in una scala da A ad F e da 1 a 6.

Metodo 4X4 EUROPOL

su una scala dei valori, sull'asse delle ascisse viene fatta una valutazione di attendibilità della fonte da 1 a 4, e stessa cosa accade sull'asse delle ordinate per la valutazione dell'informazione: laddove il risultato rientri entro il punteggio 2, l'informazione sarà da ritenersi confermata.

I codici di attendibilità della FONTE sono cosi classificati:

A= senza dubbi di autentica. Affidabile o competente oppure in passato sempre affidabile.

B= affidabile nella maggior parte dei casi.

C= non affidabile nella maggior parte dei casi.

D= non valutabilità dell'affidabilità (es. anonima).

I codici di attendibilità di valutazione/accuratezza dell'INFORMAZIONE sono:

1= sicura.

2= conosciuta personalmente dalla fonte ma non dall'agente che la riferisce.

3= non conosciuta personalmente dalla fonte ma avallata da altre informazioni già registrate.

4= non conosciuta personalmente dalla fonte e non avallabile in alcun modo

Figura 1.3: Metodo 4X4 EUROPOL[3]

Affidabilità della FONTE

A affidabile: nessun dubbio sull'autenticità, sull'affidabilità o sulla competenza della fonte. History of complete reliability. Storia di completa affidabilità

B di solito affidabile: Minori dubbi. History of mostly valid information. Storia di informazioni per lo più valide

C Abbastanza affidabile: dubbi. Provided valid information in the past. Fornito informazioni valide in passato

D Non di solito affidabile: dubbi significativi. Provided valid information in the past. Fornito informazioni valide in passato

E inaffidabile: Manca di autenticità, affidabilità e competenza. History of invalid information. Storia di informazioni non valide

F non può essere giudicato: informazioni insufficienti per valutare l'affidabilità. May or may not be reliable. Può o non può essere affidabile

Affidabilità dell'INFORMAZIONE

1 Confermato: Logico, coerente con altre informazioni rilevanti, confermato da fonti indipendenti

2 Probabilmente vero: logico, coerente con altre informazioni pertinenti, non confermato

3 Forse vero: ragionevolmente logico, concorda con alcune informazioni rilevanti, non confermate

4 Senza dubbio vero: non logico ma possibile, nessun'altra informazione sull'argomento, non confermata

5 Improbabile: non logico, contraddetto da altre informazioni pertinenti

6 Non può essere giudicato: la validità delle informazioni non può essere determinata

Figura 1.4: Griglia 6X6 NATO[3]

1.4 Strumenti e tecniche di ricerca

Seguendo i vari passaggi della metodologia di ricerca OSINT, dopo aver raccolto tutte le informazioni, si giunge al momento dell'analisi, elaborazione e filtraggio dei contenuti trovati. Nell'esecuzione delle operazioni di ricerca, strutturazione e selezione dei dati e delle informazioni, il lavoro del singolo individuo viene facilitato attraverso l'ausilio di specifiche tecnologie e strumenti avanzati, scelti a seconda dello scopo della ricerca. Infatti, le capacità di calcolo e di elaborazione dei moderni computer, unite al più recente avvento del machine learning, aiutano notevolmente durante le fasi di organizzazione e di strutturazione dei dati raccolti. Diversamente, considerata la loro più disparata natura, risulta difficile o addirittura impossibile padroneggiare enormi quantità di dati eterogenei, privi di una struttura o collegamenti tra loro, e poterne estrarre una conoscenza di qualsiasi genere.

È importante sottolineare che una ricerca completa ed esaustiva non può essere compiuta analizzando ogni singola fonte nei minimi dettagli poiché questo comporterebbe spreco di tempo raccogliendo anche informazioni inutili. Devono quindi essere scelte apposite parole chiave, tramite le quali risulti poi possibile ricostruire l'essenza del contenuto ed ottenere spunti e stimoli per nuove ricerche. Nel fare ciò, vengono quindi in aiuto all'analista o ricercatore OSINT numerosi strumenti, tecniche di ricerca e tecnologie.

Per effettuare una prima ricerca di informazioni generiche, si può semplicemente partire scrivendo il proprio input su un qualsiasi motore di ricerca come Google, Bing o Yahoo. Così facendo, utilizzando semplici query di ricerca, è possibile ottenere informazioni di diversa natura corrispondenti al nostro input. Non tutti i contenuti che vengono indicizzati sono però di fatto utili. Per questo motivo, è possibile effettuare una scrematura dei risultati utilizzando parole chiave, operatori di ricerca o filtri forniti dai browser per perfezionare le ricerche ed ottenere esattamente il tipo di informazioni di interesse.

Uno strumento di facile consultazione che può essere di aiuto per orientare meglio le ricerche successive è il framework interattivo 'OSINT framework'.

Come si può notare dalla figura sottostante, questo strumento riassume una serie di risorse prevalentemente gratuite, utili a raccogliere ulteriori informazioni di svariata natura, diversamente difficili da reperire all'interno della rete Internet. A seconda delle specifiche esigenze o dello specifico input a disposizione, questo framework interattivo consiglia una serie di strumenti che possono aiutare nel soddisfare le richieste e nel compiere le ricerche. Alcuni degli strumenti elencati potrebbero richiedere un pagamento per visualizzare ulteriori dati, ma sfruttando la versione base si dovrebbero comunque ottenere almeno una serie di informazioni rilevanti.[4]

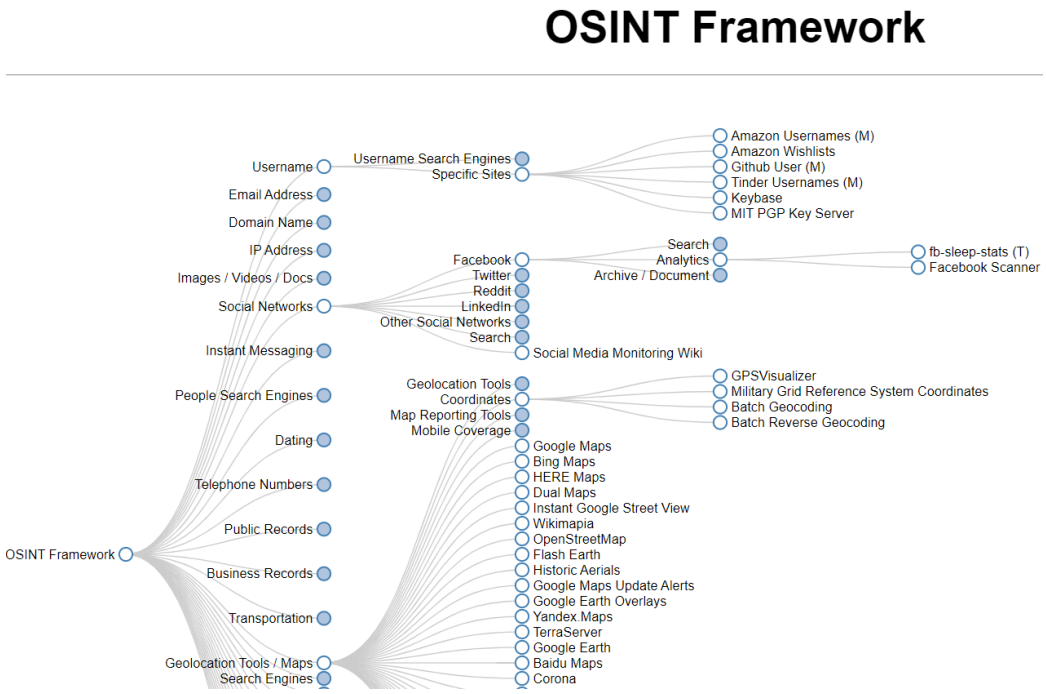


Figura 1.5: Schermata di esempio navigazione OSINT framework[4]

Analizzando più da vicino questo framework, ci si accorge di come alcuni strumenti vengano però utilizzati più frequentemente rispetto ad altri, per questioni di praticità, velocità, correttezza ma sopratutto vastità della ricerca. La seguente immagine riassume i principali strumenti di OSINT di cui fanno ampio uso gli analisti, indicando la tipologia di dati in ingresso e in

uscita, la piattaforma di funzionamento, la tipologia di interfaccia utente ed eventuali funzionalità aggiuntive.[22]

OSINT tool	Input				Output	Extensibility	Interface	Platform	Other feature
	Identity data	Network data	File data	Selectable data source					
FOCA	✗	Domain	File name, Folder	Google, Bing, DuckDuckGo	Identity info, Network info, File info	✗	Stand-alone program	Windows	Server discovery module
Maltego	Personal information, company, community	Domain	File URL	✗	Identity info, Network info, File info	Custom transforms	Stand-alone program	Linux, Windows, MAC	Location, Auto input/ output refeed, Results in oriented graph
Metagoofil	✗	Domain	File type	✗	Network info, File info	✗	Command line	Linux, Windows	Option to narrow results
Recon-NG	Personal information	Domain	✗	Several	Identity info, Network info, File info	✗	Command line	Linux	Location, Modules for discovery and exploitation
Shodan	Country, City, Keyword	Operating system, IP Address, Port, Host name	✗	✗	Network info	✗	Web interface	Online	Location, Webcam captures
Spiderfoot	Email, Real name, Phone Number	Domain, IP Address, Subnet, Host name	✗	Several	Network info	Custom modules	Web interface	Linux, Windows, MAC	Different types of scan, Results in oriented graph
The Harvester	Company	Domain, DNS server	✗	Several	Identity info, Network info	✗	Command line	Linux, Windows, MAC	Results in reports, Option to narrow files and results
IntelTechniques	Personal information, company, community	Domain, IP Address	File name, File type, File URL	Several	Identity info, Network info	✗	Web interface	Online	Location, Public records, OSINT virtual machine

Figura 1.6: Strumenti di OSINT maggiormente utilizzati[1]

Un'ulteriore distinzione tra tutti gli strumenti a disposizione può essere effettuata se si considerano le singole informazioni o i singoli dati che si devono ottenere dalla ricerca. A seguire saranno indicati i principali strumenti e le principali tecnologie utilizzate, ordinatamente suddivise per ambito o scopo di ricerca. Per ogni voce, si andrà poi ad analizzare nello specifico gli strumenti interessati, a seconda dell'input disponibile ad inizio ricerca.[1]

- *Indirizzi email*
- *Usernames*
- *Nomi propri*
- *Posizione e localizzazione*
- *Indirizzi IP*
- *Pagine Web*
- *Social Media*

1.4.1 Indirizzi email

L'uso di un indirizzo email come input di una ricerca si dimostra essere molto utile e valido quando il vero nome e cognome di una persona non è disponibile o comunque per evitare di cadere in errori come duplicati ed omonimi dell'oggetto della ricerca. È infatti dimostrato che una tecnica di ricerca che sfrutta l'indirizzo email risulta essere più veloce ed ottenga migliori risultati.

Principali strumenti utilizzati:

- *Hunter*[10]: viene utilizzato per determinare l'esistenza, correttezza e validità di un indirizzo email.
- *Have I Been Pwned*[23]: controlla se l'indirizzo email è stato soggetto di data breach, ovvero violazioni pubbliche o fughe di dati. In caso positivo, è possibile sfogliare l'elenco dei siti in cui è avvenuta tale violazione.
- *Pilp*[24]: ricerca quante più informazioni possibili circa il proprietario dell'indirizzo email in questione come nome e cognome, username, numero di telefono, istruzione, abitazione e altro.

1.4.2 Usernames

gli username sono un'importante fonte di informazione perché permettono di eseguire una ricerca trasversale in maniera automatica sulle diverse piattaforme e servizi online così come su singole pagine web.

Principali strumenti utilizzati:

- *KnowEm*[25], *Name Checkr*[26], *User Search*[27]: questi strumenti controllano l'esistenza di un determinato username sui social media, social network e sui domini più popolari.
- *NameVine*[28]: cerca sul web tutti i profili che possono corrispondere in maniera approssimata all'username dato in input, modificandolo per cercare di scovare eventuali varianti.

- *Lullar*[29]: a partire dall'username, vengono generate automaticamente degli URL appartenenti ai principali social media, senza verificare la reale esistenza del profilo. A seguire, se il collegamento creato funziona significa che il profilo esiste mentre se il link è interrotto significa che non è stata trovata alcuna corrispondenza.

1.4.3 Nomi propri

La ricerca con nome e cognome del soggetto target è la più tipica tra tutte le tipologie di ricerca. Solitamente gli strumenti impiegati riportano alla luce informazioni generiche come indirizzi di casa o lavoro, numeri di telefono, email, username e simili.

Principali strumenti utilizzati:

- *That's Them*[30], *Spokeo*[31], *Yasni*[32]: forniscono informazioni generiche come email, indirizzi residenza, lavoro, istruzione, sesso, età, paese e altre informazioni personali.

- *Geni*[33], *Family Search*[34]: questi strumenti offrono un servizio di genealogia in quanto si occupano della ricerca di informazioni sulla parentela, legami familiari ed antenati del soggetto target.

1.4.4 Posizione e localizzazione

La ricerca tramite posizioni o luoghi specifici frequentati da un soggetto può essere utile per tracciare i suoi spostamenti, conoscere le sue abitudini e preferenze. A questo scopo, sono molto utilizzati anche immagini, indirizzi fisici e coordinate GPS.

Principali strumenti utilizzati:

- *Google Maps*[9], *Wikimapia*[35], *Bing Maps*[36], *GPS Coordinates*[37]: mostrano immagini aggiornate in tempo reale della terra. Inoltre, è possibile scoprire le coordinate GPS a partire da una determinata posi-

zione così come è possibile effettuare l'operazione inversa per ottenere una determinata posizione a partire dalle coordinate GPS.

- *Land Viewer*[38], *Terra Server*[39]: mostrano immagini catturate nel tempo, visuali storiche ed obsolete di specifici luoghi.

1.4.5 Indirizzi IP e metadati

Gli indirizzi IP, insieme ai metadati, sono principalmente ottenuti tramite operazioni di ricerca su pagine web, connessioni internet o indirizzi email. Inoltre, quando ci si imbatte in un attacco di tipo informatico, sono oggetto di un'attenta analisi forense digitale per poter estrarre quante più informazioni possibili a riguardo.

Principali strumenti utilizzati:

- *IP Location*[40]: dato un indirizzo IP, si ricercano informazioni di alto livello come la posizione, il paese, la regione e la città di origine, il nome di dominio, l'ISP e altro.

- *ViewDNS*[41]: analizza l'indirizzo IP ancora più nello specifico rispetto allo strumento precedente. In particolare, offre informazioni circa il proprietario del dominio associato, eventuali domini aggiuntivi, visualizza servizi in esecuzione, analizzare reti, router e server associati.

- *I Know What You Download*[42]: monitora i torrent online con uno specifico indirizzo IP di raccolta al fine di scoprire quali file vengono scaricati dal nostro target, anche nel caso in cui si intercettino contenuti di natura personale o dati sensibili.

1.4.6 Pagine Web

Un tipico punto di interesse nelle indagini OSINT sono le pagine web poiché da esse è possibile ottenere numerose informazioni circa il nostro soggetto target.

Principali strumenti utilizzati:

- *DNS Trails*[43]: permette di estrarre record DNS insieme ad eventuali domini ad esso connessi.

- *Whoisoly*[44]: mostra risultati simili allo strumento precedente ma include ulteriori informazioni come nome del proprietario, recapiti telefonici, email e altro.

- *Wayback Machine*[45]: attraverso questo strumento di backup per siti web, è possibile monitorare ed analizzare tutte le modifiche ed i cambiamenti effettuati nel tempo.

- *Whois*[46]: offre la funzionalità di 'ping' per verificare la connettività di un sito web e la funzionalità di 'traceroute' per studiare il percorso che i dati compiono per arrivare al dominio.

- *SimilarWeb*[47], *FindSubdomains*[48]: con il primo strumento vengono mostrate statistiche relative al traffico sulla pagina web mentre con il secondo si cercano eventuali sottodomini ad esso associati.

1.4.7 SOCMINT: Social Media INTelligence

Con l'avvento dei social media e grazie alla loro continua espansione, analisti ed esperti nel mondo dell'intelligence hanno preso consapevolezza dell'utilità, importanza e rilevanza di tutti i contenuti che vi circolano all'interno.[49] Ma come estrapolare ed interpretare a proprio vantaggio tutti i dati e le informazioni disponibili? Facendo affidamento su svariate tecnologie e tecniche di ricerca avanzate come il data mining, il machine learning e la social network analysis, è possibile estrarre ed analizzare le informazioni dai social media in modo da ottenere specifici risultati a seconda dell'obiettivo stabilito. L'insieme di tutte queste attività di ricerca viene chiamato Social Media INTelligence (SMI o SOCINT).

Nell'immagine seguente vengono riportati i principali social media impiegati ed analizzati nella maggior parte delle ricerche OSINT, suddivisi ordi-

natamente per ambito o per tipologia di appartenenza come gaming, musica, video e altro.

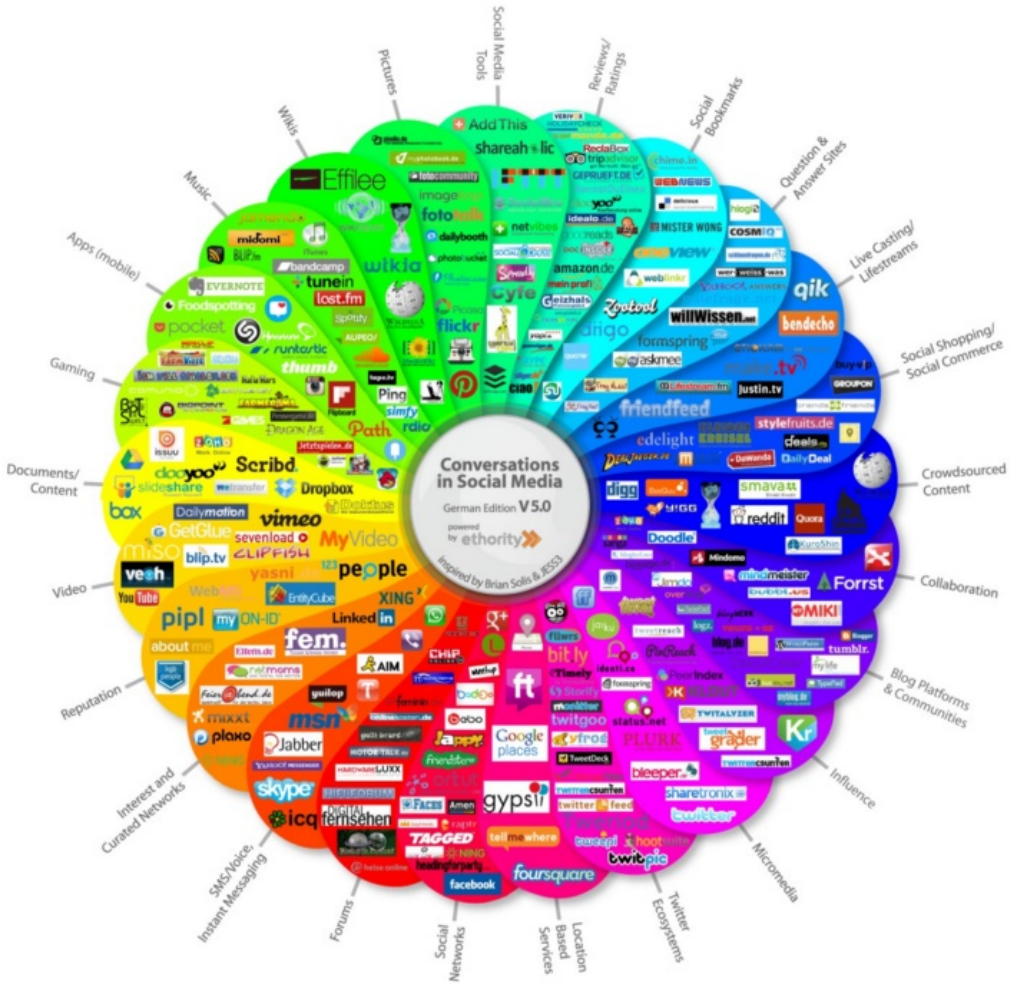


Figura 1.7: Social media per il reperimento di informazioni[3]

Strumenti e tecniche di ricerca SOCINT

Seguendo i risultati ottenuti dall'applicazione di tecniche sentiment analisys, i protocolli SOCINT si sono evoluti per andare ad insinuarsi in maniera sempre più efficiente nelle reti e nei canali social, con l'obiettivo di individuare in maniera sempre più efficace i profili e relazioni tra utenti e poter poi

costruire un profilo relazione reale, attendibile e documentabile del soggetto target. [50]

Per quanto riguarda le tecniche, gli strumenti e le tecnologie che si utilizzano durante le fasi di ricerca, non si può fare a meno di specificare come l'obiettivo da raggiungere influenzi in maniera primaria la scelta della procedura da adottare. Tra le tecniche di estrazione più comunemente utilizzate si possono citare:

- **Scraping**: tecnica con la quale si ricercano dati non strutturati contenuti in siti web di proprietà di terzi come concorrenza, forum, blog e altri. La collezione di questi dati avviene sfruttando un motore di ricerca chiamato spider unito all'utilizzo di specifici software: insieme, questi due strumenti, sfogliano ed analizzano queste pagine web alla ricerca di specifiche informazioni sulla base di parole chiave fornitegli in precedenza.

- **utilizzo di API**: le API (Application Programming Interface) sono applicazioni che espongono le funzionalità di altre applicazioni in modo da agevolarne la programmazione. Nelle fasi di sviluppo, le le API si utilizzano quindi per semplificare l'integrazione e la comunicazione tra due o più applicazioni, in modo da evitare replicazioni di codice, errori o ridondanze. [51]

- **Utilizzo di specifiche piattaforme**: software e tecnologie specifiche vengono unite per dare forma a piattaforme apposite in grado di compiere i lavori più disparati in maniera minuziosa ed efficiente. Questi software utilizzano solitamente l'intelligenza artificiale per effettuare operazioni di social listening, sentiment analysis o image recognition.

1.5 Utilizzi ed obiettivi della ricerca OSINT

Le ricerche effettuate utilizzando tecniche di OSINT hanno una vasta potenzialità che può essere sfruttata ed utilizzata nei più disparati campi: par-

tendo dalla politica e dall'economia, si arriva fino al campo della sociologia, della statistica e tante altre discipline. Considerato ciò, non è difficile pensare come le finalità, gli obiettivi e l'utilizzo delle tecniche di ricerca OSINT siano molto disparate e dipendano dall'ambito a cui si fa riferimento.

Storicamente, la ricerca OSINT veniva utilizzata solamente da enti governative, forze dell'ordine, enti statali e simili per eseguire le proprie ricerche. A questo proposito, possiamo individuare quattro macro categorie di ricerca:

- *Social opinion and sentiment analysis*: la possibilità di indirizzare uno specifico target di persone a comportarsi in un certo modo o ad andare in una certa direzione, raccogliendo dati derivanti dall'utilizzo dei social media come messaggi, preferenze, interessi ed interazioni con altri utenti.

- *Cybercrime and organisation crime*: la costante ricerca di schemi e l'analisi di dati permettono di scovare e tracciare i movimenti di organizzazioni criminali nel web, in modo da fermarli nelle prime fasi di attacco.

- *Cybersecurity and cyberdefence*: la possibilità di sfruttare le vulnerabilità e gli errori commessi in precedenza per migliorare la robustezza e la sicurezza di svariati sistemi con l'ausilio di tecniche di machine learning.

- *Open Source Intelligence per il sociale*: la collaborazione tra professionisti provenienti da tutto il mondo e da diversi campi, i quali mettono a disposizione le loro conoscenze e la loro esperienza per aiutare le istituzioni sia in attività pratiche che in attività di sensibilizzazione contro problemi come la pedopornografia, il revenge porn, la scomparsa di persone, violenze private, furto di identità e tanto altri crimini che si possono compiere sui cittadini, sia nella rete che nel mondo reale. Nel mondo del web si riuniscono sotto l'hashtag 'OsintForGood'.[50]

Altre realtà che utilizzano la metodologia di ricerca OSINT per proprie finalità si trovano spesso nell'ambito pubblico come organizzazioni internazionali.

Enti benefiche, no-profit e tante altre fanno uso di fonti derivanti da tecniche di ricerca OSINT per supportare operazioni umanitarie in tutto il mondo, ciascuna perseguendo il proprio obiettivo.

Con il passare del tempo e l'evolversi degli strumenti investigativi, gli strumenti e le tecniche di ricerca OSINT hanno iniziato a prendere piede anche in organizzazioni, aziende e compagnie private. Le grandi realtà commerciali ne fanno ampio uso in primo luogo per eseguire ricerche ed analisi di mercato, utilizzando quindi le informazioni estrapolate da ricerche OSINT per capire le strategie commerciali della concorrenza e per pianificare di conseguenza una migliore strategia commerciale. Inoltre, vengono effettuate anche analisi della reputazione del proprio brand o dei propri prodotti, in modo tale da risalire ad eventuali informazioni che possono ledere l'immagine dell'azienda ed intervenire di conseguenza per ripristinare la propria reputazione.

In ambito privato, oggigiorno si vedono anche studi legali oppure investigatori privati che utilizzano la ricerca OSINT per reperire dati, informazioni o contenuti ai quali non avrebbero avuto accesso se ricercati nelle classiche banche dati consultate abitualmente. I risultati di ricerche OSINT molto spesso finiscono per fornire informazioni ed indizi utili per ribaltare o confermare le sorti di un procedimento civile o penale.

1.6 Aspetti legali

Quando si parla di OSINT non si deve dimenticare che la ricerca e l'utilizzo di informazioni pubbliche portano con sé anche il rispetto e la protezione dei suddetti dati. Una prima garanzia di ciò si basa proprio sulla definizione di OSINT, poiché consultando e recuperando esclusivamente informazioni da fonti pubbliche, il rispetto della privacy degli individui dovrebbe essere garantito in maniera intrinseca. Spesso però, il solo fatto che le informazioni siano di natura pubblica e liberamente accessibili non significa che non siano o non contengano dati personali e sensibili. Per questo motivo, tutti i dati

raccolti in fase di ricerca devono sempre essere gestiti con cura, in maniera consapevole ed utilizzati per scopi legittimi.

Da un punto di vista legale, durante una qualsiasi attività di ricerca OSINT, la certezza di tutelare la privacy degli utenti sotto osservazione ed evitare problemi riguardanti profilazione, dispersione di informazioni o diffamazione deve sempre essere garantita. Con l'avvento del GDPR (General Data Protection Regulation) dell'Unione Europea, la normativa in materia di dati personali è cambiata: la definizione di dato personale è stata estesa e, sempre in questo senso, se diverse informazioni raccolte insieme possono portare ad identificare una persona, allora vengono considerate come dato personale e sensibile, anche se crittografate o anonime. Chiunque effettui ricerche deve quindi essere in grado di svolgere le proprie investigazioni nel rispetto del trattamento del dati personali come regolamentato dal GDPR. [52]

Nell'ampio mondo dell'OSINT, a chi è quindi consentito per legge effettuare ricerche e chi ha il diritto di curiosare o accedere a tutti i dati privati e le informazioni più personali dei singoli cittadini? Come si può proteggere i termini giuridici sia il ricercatore che il ricercato?

Ai sensi dell'art. 134 TULPS, ovvero del Testo Unico delle Leggi di Pubblica Sicurezza, chiunque effettui ricerche per conto di privati deve ottenere una licenza apposita, in quanto *"senza licenza del prefetto è vietato ad enti o privati di prestare opere di vigilanza o custodia di proprietà mobiliari od immobiliari e di eseguire investigazioni o ricerche o di raccogliere informazioni per conto di privati."* [53] Nel caso in cui avvenga una violazione di quanto appena scritto, il rischio è quello di incorrere in un reato penale, per mezzo del quale *"I contravventori saranno puniti con l'arresto fino a due anni e con l'ammenda da euro 206 ad euro. 619."*[54]

Come ormai si ha intuito, la metodologia di ricerca OSINT è davvero potente e risulta quindi essere pericolosa se utilizzata in maniera impropria. È però facile immaginare come nel vasto mondo dell'OSINT non sempre si operi in maniera corretta e lecita. Se da un lato il suo impiego viene

correttamente giustificato ed impiegato per scopi legittimi, dall'altro l'utente potrebbe essere un delinquente che tenta di commettere un crimine. In questo contesto, si possono citare una serie di casistiche come:

- *organizzazioni terroristiche*: utilizzano le fonti OSINT per pianificare gli attentanti e muovere al meglio la loro organizzazione
- *ladri, estorsionisti e altri malintenzionati*: è possibile analizzare i membri della famiglia per rubare da casa nel momento migliore oppure pubblicare informazioni private e personali della vittima in cambio di riscatto.
- *hacker o altri professionisti informatici*: tecniche OSINT vengono impiegate per raggiungere una serie di obiettivi illeciti o per compiere azioni come cyberbullismo, cybergossip o cyber-vittimizzazione.
- *altri intrusi*: persone che per passione o per ricerche personali riescono ad entrare in possesso di informazioni di questo tipo eseguendo operazioni di tipo OSINT.

A fronte di ciò, risulta essere di primaria importanza controllare che tutti gli strumenti, le tecnologie e le tecniche di ricerca impiegate siano conformi alle normative e vengano utilizzati correttamente senza ledere diritti o libertà altrui. Nello svolgere questo compito, potrebbe essere utile distinguere i diversi ruoli degli utenti che effettuano le ricerche tramite OSINT ed attribuire a ciascuna categoria i relativi privilegi, in modo da limitare l'accesso alle risorse.[52]

- Investigatori privati, dipendenti e persone che lavorano in ambito privato hanno accesso alle informazioni di base e a strumenti di ricerca non troppo invasivi.
- Governi e forze dell'ordine possono invece indagare utilizzando tutti gli strumenti a loro disposizione e hanno il completo accesso a tutte le informazioni reperite.

In conclusione, affinché le attività di OSINT siano conformi alle regola e alla normativa vigente a protezione dei dati personali dei soggetti target, i ricercatori devono verificare la legittimità e la qualità delle fonti, minimizzare i dati raccolti, cancellare quelli non necessari ai fini della ricerca, proteggere e garantire i diritti ai titolari degli stessi. Inoltre, è vietato loro comunicare a terzi tutti i risultati trovati. [55]

1.7 Esempio di progetti

Riprendendo l'obiettivo dell'elaborato di tesi, ovvero lo sviluppo e l'integrazione tra i principali strumenti utilizzati dalla ricerca OSINT, risulta difficile trovare nel web un'applicazione che riunisca e faccia collaborare tra loro in maniera logica e strutturata altri strumenti con l'obiettivo di ottenere risultati accurati ed attendibili da una ricerca.

Questa situazione può essere da un lato giustificata in quanto come ormai si avrà capito, la ricerca OSINT va intesa come un intero processo di analisi, ricerca ed elaborazione e non solo come l'utilizzo in successione dei singoli strumenti. Mappe, schemi e workflows risultano essere fondamentali per riordinare tutti i dati ottenuti e l'intervento umano è comunque indispensabile per eliminare tutti quei dati "raw" che vengono automaticamente generati dalle ricerche, in modo da ottenere ricerche accurate e precise.

Nonostante queste prime considerazioni, non si può però fare a meno di notare come lo sviluppo di un'integrazione tra gli strumenti possa velocizzare e migliorare il processo di ricerca ed analisi, il tutto se svolto in maniera accurata e studiata. L'intervento umano resta comunque un elemento alla base del processo di ricerca ma la tecnologia viene in aiuto in numerosi fasi, principalmente durante l'esplorazione, la raccolta e l'elaborazione dei dati e delle informazioni trovate. Questa è la motivazione cardine sulla quale è stato pensato e sviluppato l'elaborato di tesi che nei capitoli successivi si andrà a presentare in maniera dettagliata.

A sostegno dell'utilità di questo progetto, previa una più attenta e dettagliata ricerca nel web, è stato possibile trovare alcuni strumenti che svolgono il compito di automatizzazione di ricerca OSINT. A seguire si indicheranno gli strumenti di maggiore interesse trovati, ciascuno presentato con una semplice analisi di funzionalità e caratteristiche generali, cercando di visualizzare e raccogliere idee o miglioramenti per sviluppi ed integrazioni future.

- Creepy[5]
- Maltego[6]
- OSINT Integration Tool[7]

1.7.1 Creepy

Creepy è uno strumento di ricerca OSINT concentrato unicamente sulla funzionalità di geolocalizzazione di un determinato target. Utilizzando questa tecnologia è possibile raccogliere informazioni riguardanti il luogo e la posizione di una determinata fonte o informazione trovata nella rete Internet.

In maniera più specifica, è possibile raccogliere informazioni partendo da posizioni fisiche, coordinate geografiche oppure da posizioni rilevate tramite social media: risulta così evidente come più strumenti possano essere utilizzati in input in modo da ottenere successivamente un unico risultato in output.

In aggiunta, è possibile affinare la ricerca filtrando i risultati della ricerca in base alla posizione o alla data esatta e visualizzare i risultati su una mappa. Tutte le informazioni possono poi essere raccolte in un file formato csv oppure KLM per permetterne l'esportazione ed essere sottoposte ad un'ulteriore analisi utilizzando lo strumento Google Maps.

Nella seguente immagine si può vedere come appare all'utente la schermata di ricerca dello strumento: sulla sfondo è presente una mappa di orientamento generale mentre nel punto indicato dal segnaposto rosso, ovvero quello di nostro interesse, ci viene mostrata una foto per maggiore chiarezza.

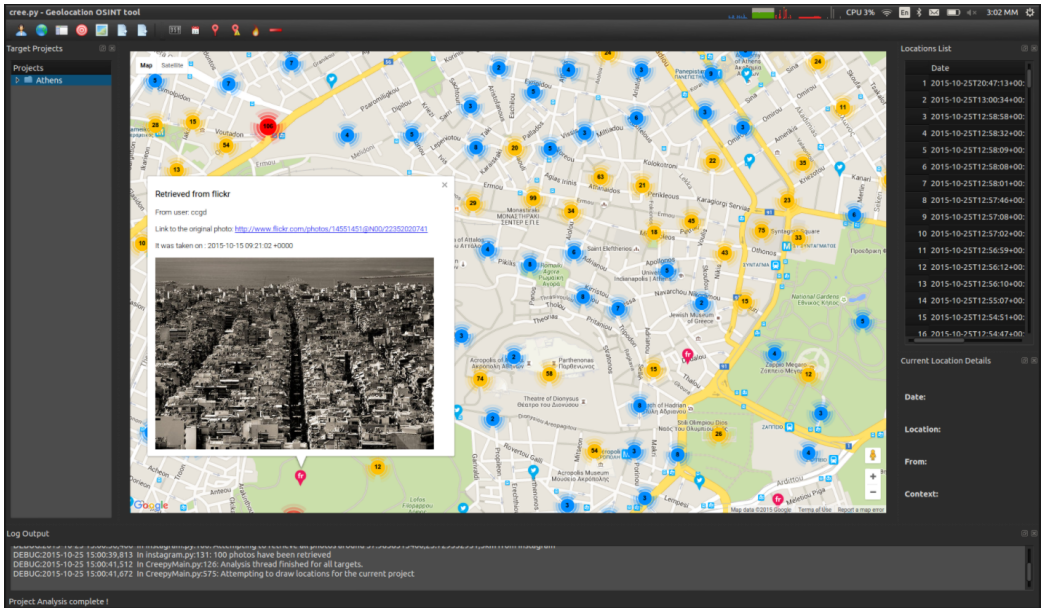


Figura 1.8: Schermata esempio di Creepy[5]

1.7.2 Maltego

Maltego è uno strumento utilizzato nelle ricerche OSINT per la sua funzionalità di raggruppamento e visualizzazione delle informazioni. Infatti, a differenza delle tecnologie analizzate fino ad ora che supportano in maniera attiva le fasi di analisi e ricerca, questo strumento offre la possibilità di effettuare grafici e collegamenti tra tutti i dati fornitogli in input.

Si può quindi considerare come uno strumento completo che aiuta l'analista nella fase di raccolta e organizzazione dei dati, in quanto attraverso operazioni di data mining ed analisi in tempo reale di dati, rappresenta poi queste informazioni su un grafico basato su nodi, facilitando l'identificazione di connessioni tra tutti i dati raccolti dalla ricerca.

Utilizzando Maltego risulta quindi facile estrarre dati da fonti di ricerca OSINT, unire automaticamente tutte queste informazioni in modo da formare un grafico comprensibile e successivamente mapparle visivamente per poter esplorare tutto il panorama di informazioni.

Per ottenere risultati da una ricerca eseguita tramite Maltego, non si uti-

lizzano in input informazioni provenienti unicamente da ricerche OSINT ma si integrano in maniera molto facile anche strumenti di terze parti utilizzando il Maltego Transform Hub. Tra gli strumenti più comunemente integrati e tra quelli citati in precedenza si possono trovare Pilp, Google Maps, Google, Have I Been Pwned, Shodan, Wayback Machine, TinEye e tanti altri.

La seguente immagine mostra una schermata di utilizzo dello strumento, visualizzando le informazioni trovate e tutti i collegamenti tra di esse in formato grafico. Più nello specifico, nella colonna di sinistra si trova la legenda di ciascuna figura, in altro viene mostrato il menù delle opzioni disponibili mentre nella colonna di destra è possibile analizzare più dettagliatamente le singole informazioni.

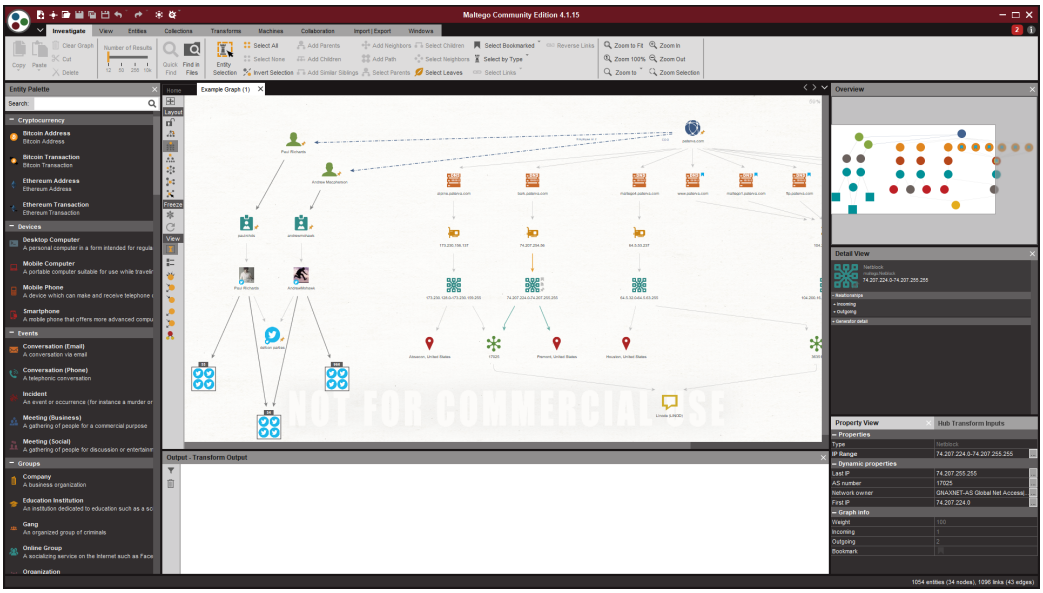


Figura 1.9: Schermata esempio di Maltego[6]

1.7.3 Sandian OSINT Integration Tool

L'obiettivo di questo strumento è quello di rilevare all'interno dei servizi di posta tutte quelle email di spam o phishing, facilitando il loro riconoscimento e aumentando il grado di protezione nei confronti degli utenti. La difficoltà

del riconoscimento sta proprio nelle email stesse, le quali vengono create in maniera sempre più personalizzata e sempre più simili alle originali.

Per scovare le eventuali email dannose, lo strumento in questione fa uso di un sistema di identificazioni di minacce sia da parte di persone fisiche che da un processo automatizzato anti-spam OSINT, segnalando come email dannose tutte le eventuali corrispondenze che si vengono a creare confrontando questi dati con alcuni già noti e registrati in appositi database.

Nella seguente immagine viene illustrato in maniera più dettagliata il flusso di lavoro che questo strumento utilizza nell'effettuare le operazioni di ricerca, controllo ed analisi, immaginando di trovarsi nello scenario in cui queste email vengano recapitate nella casella di posta di un utente.

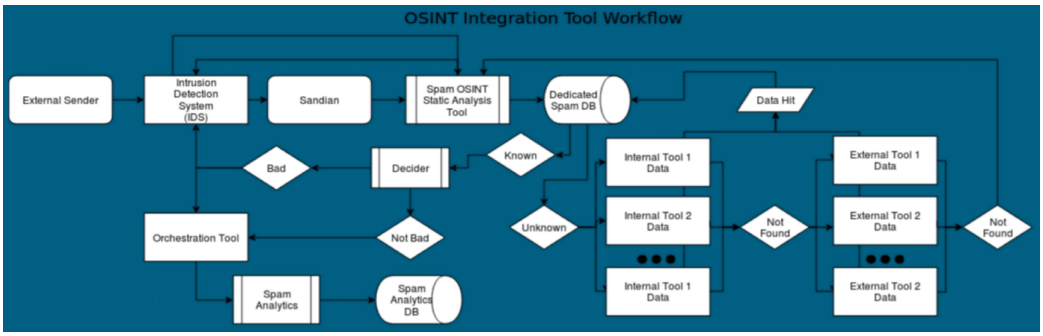


Figura 1.10: Flusso di lavoro Sandian OSINT Integration Tool[7]

Capitolo 2

Strumenti e tecnologie utilizzate

All'interno di questo secondo capitolo si andranno ad analizzare nel dettaglio i linguaggi di programmazione, gli strumenti, le tecnologie e l'ambiente di lavoro utilizzati per lo sviluppo dell'elaborato di progetto. Il lavoro svolto è stato realizzato principalmente in tre fasi: analisi degli strumenti a disposizione, studio e progettazione di un workflow di ricerca, implementazione del workflow sulla base degli strumenti selezionati.

Nella prima fase di analisi si è cercato di capire non solo quali fossero le funzionalità dei singoli strumenti di terze parti impiegati nel progetto ma anche quali fossero i loro input e output di ricerca, in modo da poter già pensare ad una possibile interazione ed integrazione tra di essi.

Sulla base dei risultati precedentemente trovati, nella seconda fase ci si è concentrati sulla progettazione di un diagramma di lavoro che potesse riportare le varie fasi di ricerca utilizzando il software online LucidChart.

Infine, nella terza e più importante fase, ha avuto inizio la vera e propria implementazione del progetto. Qui verrà descritto il sistema operativo Windows, utilizzato per lo sviluppo del software, insieme alla descrizione di GitLab e Git per Windows: il primo è la piattaforma web che ha permesso la gestione del repository del progetto mentre il secondo ha permesso

l'emulazione della shell Bash per eseguire i comandi in ambiente Windows. A seguire verrà presentato Spyder, ovvero l'IDE utilizzato per lo sviluppo del progetto. Per ultimo verranno introdotti Python, ovvero il linguaggio di programmazione utilizzato per l'implementazione delle principali funzionalità dell'elaborato e script Bash, impiegati per l'integrazione di altre funzionalità.

2.1 Strumenti impiegati

Considerando le prestazioni, i risultati e gli input/output di ciascuno degli strumenti citati nel capitolo 1, lo sviluppo dell'elaborato di tesi è stato pensato sulla base degli strumenti che ora si andranno a presentare.

2.1.1 Crosslinked

Crosslinked[56] è uno strumento di ricerca OSINT che raccoglie nomi di dipendenti appartenenti ad un'organizzazione target, sfruttando tutte le informazioni accessibili dai motori di ricerca sulla piattaforma LinkedIn.

Perchè proprio LinkedIn? Questa piattaforma di networking è una tra le più grandi ed importanti a livello professionale e viene utilizzata ogni giorno da piccole e grandi organizzazioni per il reclutamento, il marketing e la connessione interna. Come spesso accade in questi contesti, ci si trova davanti ad un'arma a doppio taglio: se da un lato una completa e dettagliata descrizione del profilo aziendale può evidenziare i punti di forza dell'azienda agli occhi di terzi, dall'altro l'esposizione online di queste informazioni risulta essere un'ottima fonte per il reperimento di dati riguardanti i dipendenti, la loro posizione e l'azienda nel complesso.

CrossLinked è appositamente ideato per semplificare il processo di ricerca su questa piattaforma. Tutte le operazioni vengono infatti eseguite senza l'utilizzo di chiavi API, credenziali di accesso e senza interagire direttamente con il sito.

Al loro posto vengono invece utilizzate apposite query di ricerca che scavano nel web per riportare alla luce tutti i risultati corrispondenti ad un

determinato formato, specificato nella riga di comando ad inizio ricerca. I nomi utente trovati possono poi essere utilizzati per ottenere ulteriori dati come indirizzi email, account di dominio e tanto altro. I principali formati di ricerca utilizzati sono i seguenti:

```
crosslinked.py -f '{first}.{last}@company.com' 'Company'
```

```
crosslinked.py -f 'domain\{f}{last}' 'Test Company'
```

```
crosslinked.py -f '{first}{l}@xyz.com' 'Org XYZ'
```

Prima di poter applicare CrossLinked all'interno di un flusso di ricerca, è buona norma trovare quale convenzione utilizzi l'organizzazione in questione per la denominazione degli account utente. Questo permetterebbe di velocizzare la ricerca, poiché il formato della query di ricerca da adottare sarebbe già noto.

La seguente immagine mostra un semplice scenario di utilizzo di questo strumento: nella prima riga si trova il formato utilizzato per la query di ricerca mentre nelle righe successive vengono visualizzati i risultati. Una volta completata l'esecuzione, i nomi vengono controllati per eliminare eventuali duplicati e poi vengono scritti in un file names.txt nella directory corrente.

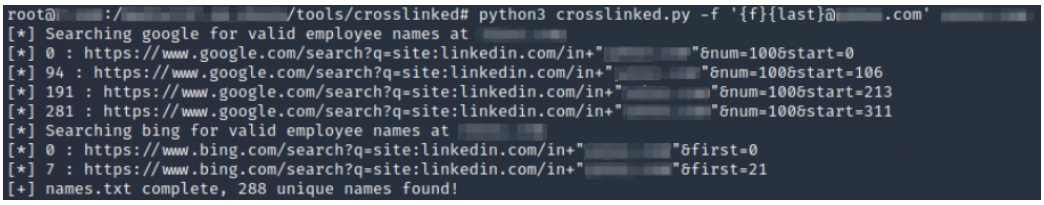


Figura 2.1: Schermata esempio per l'utilizzo di CrossLinked [8]

2.1.2 GoogleMaps

Google Maps[9] è un servizio geografico gratuito che permette di ricercare e visualizzare carte geografiche utilizzando semplicemente una connessione

Internet, accedendo da un qualsiasi browser web e utilizzando un qualsiasi dispositivo come PC, smartphone, tablet, smartwatch e persino auto.

Google Maps mette a disposizione diverse tipologie di mappe, ciascuna presentata con proprie caratteristiche e funzionalità[57]. Si andranno ora ad analizzare:

- **Visione Mappa**: utilizzata come versione di default all'apertura del servizio da browser web o da app mobile, l'interfaccia mostra una semplice cartina topografica. Seguendo il livello di zoom desiderato, vengono mostrate più o meno informazioni dettagliate circa i luoghi di interesse (indirizzi ma anche attività come ristoranti, hotel, bar, supermercati, musei e altre attrazioni turistiche), il tutto nella zona della mappa in cui ci si trova.

- **Google Maps Satellite**: in qualsiasi momento della navigazione è possibile passare a questa interfaccia, nella quale vengono mostrate le immagini satellitari di una mappa. Queste immagini però non vengono aggiornate periodicamente e quindi non sono affidabili per ottenere quelle informazioni in tempo reale. Questo servizio è invece offerto dalla versione navigatore.

- **Google Maps Navigatore**: l'interfaccia di questa opzione mostra in maniera chiara e semplice il percorso e la direzione da seguire per raggiungere un determinato luogo. Le indicazioni vengono fornite a piedi, in auto o con i mezzi pubblici e una voce aiuta nell'orientamento durante tutto il percorso.

- **Visuale Street View**: in questa interfaccia viene mostrata la strada usando immagini reali a 360 gradi, trasmettendo quindi la sensazione di trovarsi fisicamente nel luogo ricercato. Purtroppo non tutte le zone potrebbe essere coperte da questa tipologia di visione a causa di svariate difficoltà nel reperimento delle immagini.

Tra tutte le potenzialità, funzionalità ed integrazioni offerte da Google Maps, nello sviluppo di questo progetto di tesi è stato impiegato lo strumento di geocodifica.

Con geocodifica[58] si intende il processo di conversione degli indirizzi fisici (via, numero civico, città, CAP, stato) in coordinate geografiche (latitudine e longitudine) mentre con geocodifica inversa si intende il processo di conversione inversa che, a partire da determinate coordinate geografiche, riporta l'indirizzo fisico corrispondente.

Per lo svolgimento di queste operazioni, Google Maps mette a disposizione apposite API da integrare nei propri progetti utilizzando semplici richieste HTTP, in modo tale da semplificare la ricerca di un determinato indirizzo e posizionare mappe o indicatori su di esso.

La seguente figura mostra come sia possibile sfruttare questo servizio utilizzando l'API JavaScript fornita direttamente da Google Maps, accedendo tramite browser web.

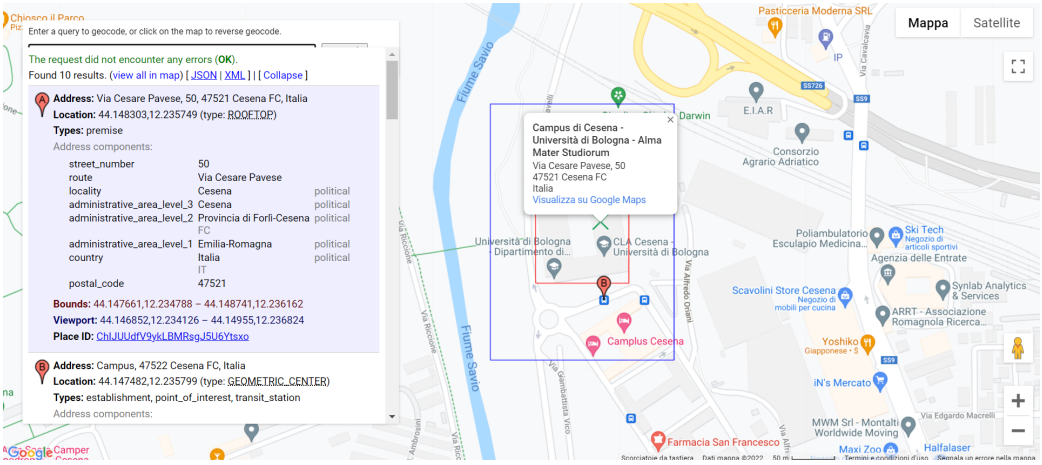


Figura 2.2: Funzionalità di reverse geocode tramite l'API JavaScript di Maps[9]

Il box in alto a sinistra permette l'inserimento delle coordinate geografiche che verranno geocodificate mentre nel pannello sottostante vengono mostrate tutte le corrispondenze trovate dopo aver avviato la ricerca. Sullo sfondo

viene semplicemente riportata la cartina topografica sulla quale si trovano inseriti in rosso tutti i segnaposto, corrispondenti ai risultati della ricerca.

Gli input di ricerca possono essere forniti sia in maniera statica, ovvero conoscendo a priori l'oggetto della nostra ricerca, sia in maniera dinamica, ovvero tutti quei casi in cui è richiesto l'input da parte di un utente. Per snellire le varie operazioni di ricerca degli indirizzi, quando possibile, si possono caricare file in formato .csv ed eseguire le ricerche su più indirizzi contemporaneamente.

Nella restituzione dei risultati di geocodifica inversa, si cerca di trovare la posizione richiesta entro una certa tolleranza. Per questo motivo vengono riportati più indirizzi associati alle coordinate, partendo dall'indirizzo con una minore corrispondenza fino a quello con corrispondenza maggiore, che di solito risulta essere quello esatto.

2.1.3 Hunter

Hunter.io[10] è uno strumento di ricerca OSINT utile per la ricerca ed il controllo della validità di indirizzi email e nomi di dominio, appartenenti a persone fisiche o ad aziende.

Per spiegare al meglio le sue principali funzionalità si osservi la seguente figura, nella quale vengono mostrati i principali servizi forniti da Hunter.

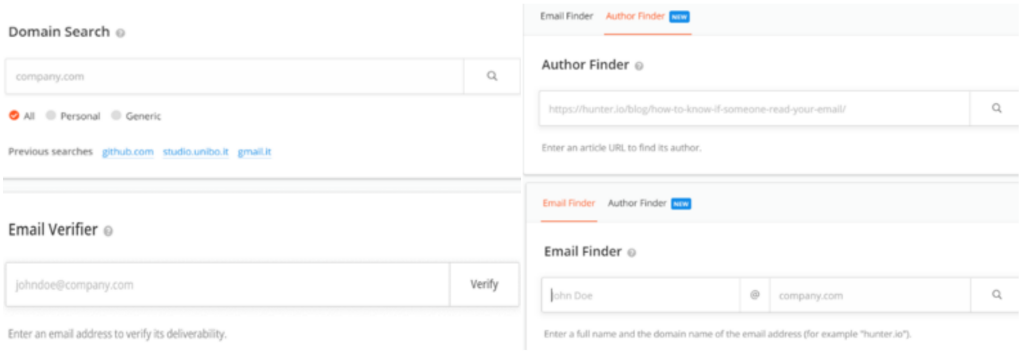


Figura 2.3: Funzionalità di ricerca e controllo validità fornite da Hunter.io [10]

- **Domain Search**: utilizzando come input di ricerca un nome di dominio, in output vengono restituiti molteplici risultati come indirizzi e-mail personali o basati sui ruoli, formati di posta elettronica, elenchi e recapiti dei dipendenti, insieme ad altre informazioni relative sempre all'azienda target.
  Inoltre, è possibile consultare le fonti utilizzate in fase di ricerca, ricevere punteggi circa l'affidabilità dei risultati trovati e salvare gli stessi nel formato desiderato.

- **Email Finder**: attraverso un nome di persona e un nome di dominio ad essa associato, è possibile ricevere in output l'indirizzo email professionale o privato del soggetto target. Vengono inoltre restituite informazioni circa l'affidabilità di questa ricerca e la fonte di provenienza delle informazioni trovate.

- **Author Finder**: viene richiesto l'URL di articolo online in input e come output vengono fornite informazioni circa il nome dell'autore, il suo indirizzo email, il punteggio di affidabilità e la fonte di provenienza.

- **Email Verifier**: dopo aver fornito un indirizzo email come input di ricerca, Email Verifier ne esegue un controllo completo restituendo informazioni circa la consegna, l'affidabilità e la reperibilità dell'indirizzo email nella rete.

In alternativa a quanto appena descritto, Hunter mette a disposizione ulteriori modalità per la fruizione dei propri servizi:

- *Estensione Google*: l'estensione Chrome di Hunter consente di trovare indirizzi email associati ad un articolo, ad una persona specifica o fonti pubbliche a partire dalla pagina in cui ci si trova nella navigazione.

- *Google Sheets*: è possibile ricercare un dominio e verificare la validità di indirizzi appartenenti ad un determinato target direttamente dal foglio Google.

- *chiavi API*: utilizzando le chiavi API, integrabili ovunque esse servano, è possibile distribuire i servizi in maniera più rapida, semplice ed efficiente. A ciascuna funzionalità è associata la propria API. In questo modo, trovare indirizzi email da nomi a dominio, ricercare informazioni sull'autore di un qualsiasi articolo oppure ottenere scansioni e risultati completi circa un indirizzo email risulta molto più facile.

  A seguito si elencano le API associate ai principali servizi proposti da Hunter:

```
GET https://api.hunter.io/v2/domain-search?
    domain=intercom.io&api_key=API_KEY (Domain
    Search)
```

```
GET https://api.hunter.io/v2/email-finder?
    domain=reddit.com&first_name=Alexis&
    last_name=Ohanian&api_key=API_KEY (Email
    Finder)
```

```
GET https://api.hunter.io/v2/author-finder?url=
    https://hunter.io/blog/how-to-know-if-
    someone-read-your-email/&api_key=API_KEY (
    Author Finder)
```

```
GET https://api.hunter.io/v2/email-verifier?
    email=patrick@stripe.com&api_key=API_KEY (
    Email Verifier)
```

- *Ricerca a blocchi*: nel caso in cui si possedessero file contenenti elenchi di nomi di persone, indirizzi email, nomi di dominio o indirizzi web di articoli, è possibile effettuarne il caricamento utilizzando una sezione apposita, in modo tale da velocizzare il processi eseguendo l'operazione su tutti gli elementi contemporaneamente.

2.1.4 Sherlock

Sherlock[59] rappresenta un potente tool di ricerca OSINT in quanto inserendo come input un semplice username, vengono ricercati nel web possibili altri account che utilizzano lo stesso username, e quindi riconducibili allo stesso utente.

L'elenco di tutti i siti e di tutti i social network che sono interessati dalla ricerca tramite Sherlock è consultabile al seguente indirizzo:`https://github.com/sherlock-project/sherlock/blob/master/sites.md`.

Dopo aver quindi impostato tutti i parametri di ricerca che possono coinvolgere uno o più username, Sherlock restituirà come output diversi file di testo quanti sono gli username, ciascuno contenente l'elenco di tutti gli account con i quali è stata trovata una corrispondenza. All'interno di questi file sono presenti anche i relativi collegamenti URL agli account trovati, per poter accedere direttamente alle informazioni cliccando sul link.

Considerando che spesso gli account social non tutelano la privacy dei propri utenti e considerando anche la noncuranza di questo aspetto da parte degli utenti stessi, spesso ci si imbatte in dati personali come indirizzi email, indirizzi di residenza o domicilio, foto di ogni tipo, titolo di studio, data di nascita, amici, familiari e tanto altro.

Nella seguente immagine si può vedere come vengono visualizzati i risultati della ricerca effettuati tramite Sherlock. La schermata di sinistra mostra il contenuto del terminale (lo stesso dal quale viene avviata la ricerca) mentre al centro viene mostrato il file di testo contente gli URL trovati, visibili anche dal terminale, ma disposti in maniera più ordinata. La schermata a destra mostra semplicemente la directory in cui si trova il file di testo.

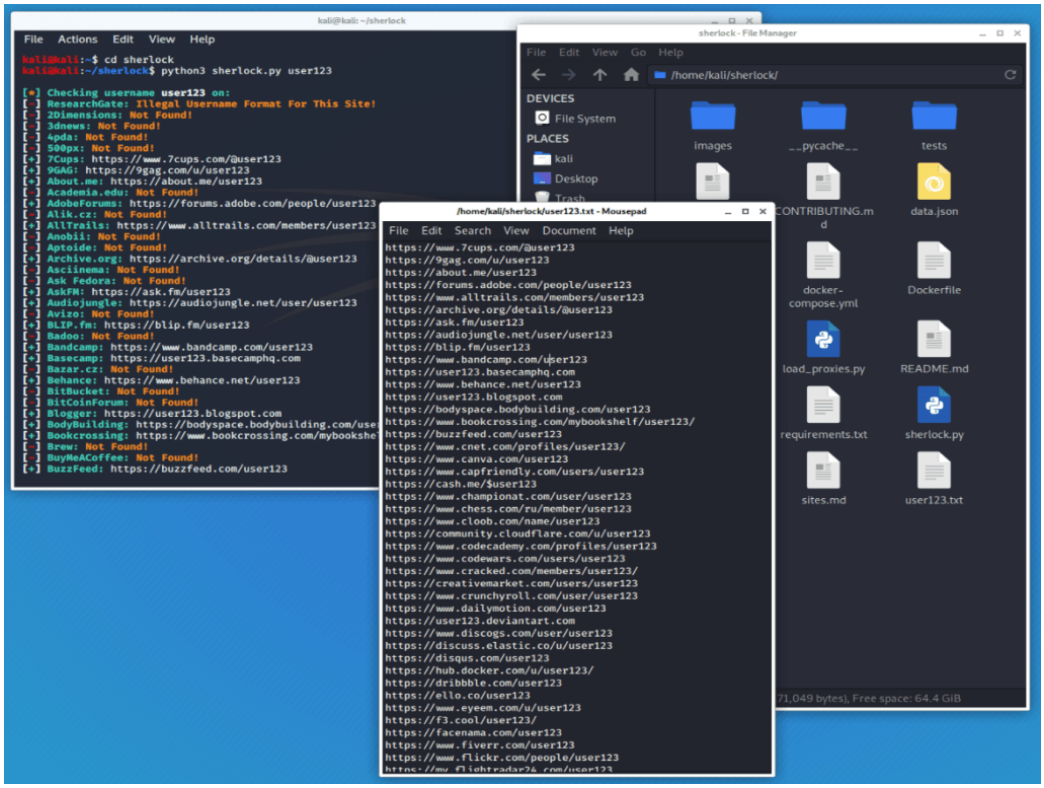


Figura 2.4: Visualizzazione risultati della ricerca tramite Sherlock[11]

2.1.5 Spiderfoot

Spiderfoot[60] è un ulteriore strumento di ricerca OSINT che si pone l'obiettivo di raccogliere nel web quante più informazioni possibili relative ad un utente, cercando di estrarre valore informativo da ogni piccola traccia.

Si tratta di uno strumento versatile ed intelligente, poiché a seconda dell'input fornitogli seleziona automaticamente i moduli da attivare, in modo ottimizzare e rendere più efficienti le ricerche. In alternativa, i moduli possono essere abilitati dall'utente a seconda delle proprie necessità. In aggiunta, l'utente può anche scegliere il livello di ricerca tra quattro tipologie di scansioni:

- *Passive*: vengono raccolte quante più informazioni possibili evitando rapporti diretti con il sito o gli account in possesso dal target, in modo

tale da proteggere la propria identità investigativa.

- *Investigate*: vengono effettuate una serie di scansioni circa la vulnerabilità e la pericolosità del target.

- *Footprint*: si identifica la topologia di rete del target e vengono raccolte generiche informazioni a partire dal web e dai motori di ricerca, utili ad eseguire basiche operazioni di indagine.

- *All*: consigliabile per quando si necessita di informazioni dettagliate, vengono consultate tutte le possibili risorse e fonti d'informazione disponibili, direttamente o indirettamente collegabili al target. Per questo motivo, i tempi di elaborazione possono essere lunghi.

Una volta specificato il target di ricerca insieme a tutti i relativi parametri, Spiderfoot inizia la sua indagine nel web, raccogliendo quante più informazioni possibili. La natura di queste informazioni è molto disparata: indirizzi IP, nomi di dominio, indirizzi e-mail, numeri di telefono, nomi reali, nomi host, sottoreti di rete, ASN e altro.

Al termine della ricerca, i risultati vengono mostrati tramite un semplice elenco oppure ordinatamente rappresentati in un grafico di nodi, insieme a tutte le entità, i collegamenti e le relazioni trovate tra di essi.

Tutte le potenzialità, gli impieghi e le principali funzionalità di questo strumento possono essere riassunte nella seguente lista:

- *Elaborazione dei dati*: l'obiettivo primario di questo strumento è quello di estrarre quante più informazioni possibili da ciascuna scansione. Per questo motivo, ogni dato raccolto viene elaborato da più moduli, in modo da estrarre quanto più valore possibile.

- *Semplicità e Velocità*: grazie all'utilizzo di una semplice interfaccia utente accessibile da qualsiasi browser web, Spiderfoot risulta essere intuibile e di facile utilizzo. L'utilizzo di questa interfaccia multi-target permette quindi di velocizzare sia l'avvio delle scansioni che la navigazione tra i risultati delle ricerche.

- *Portabilità*: pensato per essere ospitato e gestito nel cloud, Spiderfoot non richiede nessuna installazione ed è sempre disponibile, non creando problemi di dipendenza o compatibilità.

- *Dark web*: l'integrazione con la rete TOR permette di scansionare in maniera anonima e sicura anche i siti contenuti nel dark web.

- *API*: come alternativa di utilizzo all'interfaccia web, Spiderfoot rilascia una serie di API utili per eseguire scansioni, interrogare dati e svolgere altre operazioni direttamente nel codice di un qualsiasi progetto.

Nonostante in questo progetto di tesi si sia scelto di integrare Spiderfoot utilizzando le API, inserendo la seguente figura si vuole illustrare in maniera semplice e diretta come appaiono i dati al termine di una scansione, utilizzando l'interfaccia grafica disponibile dal browser web.

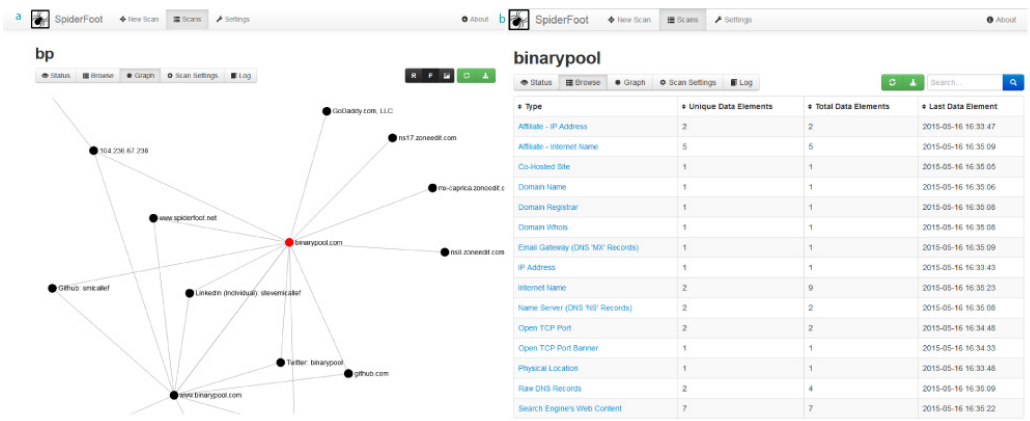


Figura 2.5: Schermate esempio per l'utilizzo di Spiderfoot tramite browser web[12]

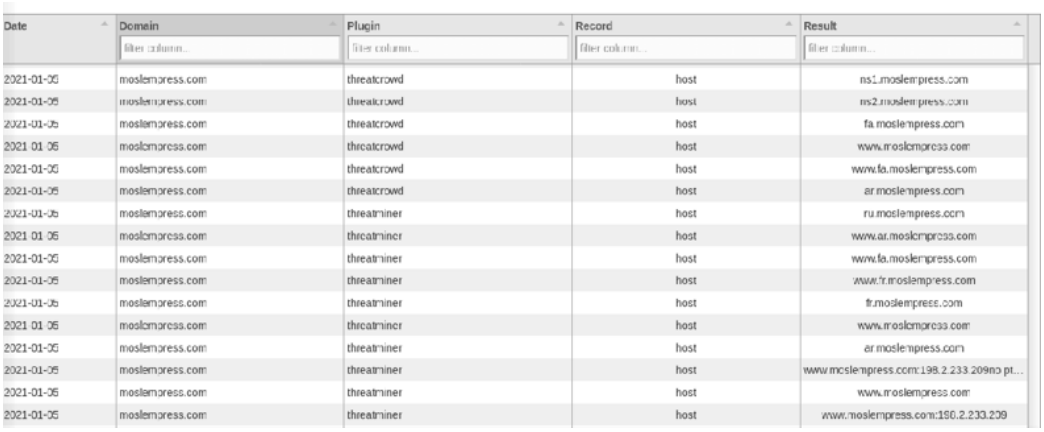
Nella schermata di sinistra vengono visualizzati attraverso un grafico tutte le informazioni e tutti i collegamenti tra di esse; nella schermata di destra, invece, le stesse informazioni vengono disposte in un semplice elenco.

2.1.6 the Harvester

theHarvester[61] è l'ultimo tra gli strumenti di ricerca OSINT analizzati ed integrati in questo progetto di tesi. Sviluppato in Python, questo strumento consente la raccolta di informazioni come indirizzi email, nomi utente, nomi host, sottodomini, indirizzi IP e URL, richiedendo come unico dato in ingresso un nome di dominio.

Le principali fonti di ricerca che vengono consultate sono i motori di ricerca ed i social come Baidu, Bing, Qwant, Google, Twitter, Linkedin e Trello, ma ci si appoggia anche a strumenti di terze parti come Shodan, dnsdumpster, Duckduckgo, Hunter e SecurityTrails (ricerca passiva). In aggiunta, vengono ricercati anche sootodomini, email referenziali e vengono applicate tecniche di reverse DNS (ricerca attiva).

Per accedere in maniera facile ed intuitiva ai risultati della ricerca, theHarvester ne propone una rappresentazione in formato .html, come mostrato nella seguente figura.



Date	Domain	Plugin	Record	Result
2021-01-05	moslempress.com	threatcrowd	host	ns1.moslempress.com
2021-01-05	moslempress.com	threatcrowd	host	ns2.moslempress.com
2021-01-05	moslempress.com	threatcrowd	host	fa.moslempress.com
2021-01-05	moslempress.com	threatcrowd	host	www.moslempress.com
2021-01-05	moslempress.com	threatcrowd	host	www.fa.moslempress.com
2021-01-05	moslempress.com	threatcrowd	host	ar.moslempress.com
2021-01-05	moslempress.com	threatminer	host	ru.moslempress.com
2021-01-05	moslempress.com	threatminer	host	www.ar.moslempress.com
2021-01-05	moslempress.com	threatminer	host	www.fa.moslempress.com
2021-01-05	moslempress.com	threatminer	host	www.fr.moslempress.com
2021-01-05	moslempress.com	threatminer	host	fr.moslempress.com
2021-01-05	moslempress.com	threatminer	host	www.moslempress.com
2021-01-05	moslempress.com	threatminer	host	ar.moslempress.com
2021-01-05	moslempress.com	threatminer	host	www.moslempress.com:198.2.233.209no pt...
2021-01-05	moslempress.com	threatminer	host	www.moslempress.com
2021-01-05	moslempress.com	threatminer	host	www.moslempress.com:198.2.233.209

Figura 2.6: Risultati di una ricerca theHarvester aperta nel browser web [13]

La tabella riportata sopra mostra diverse opzioni per il filtraggio dei risultati, in modo da ricondurre alla fonte di provenienza. Inoltre, anche se non riportato nell'immagine, vengono creati svariati grafici che rappresentano i risultati sotto diversi punti di vista. Per ultimo, viene mostrato un riepilogo di tutti i risultati ottenuti da ciascuna sorgente interrogata.

2.2 Progettazione workflow

Dopo una prima fase di analisi e comparazione tra i vari strumenti elencati, si è cercato di mettere insieme in maniera sempre più curata e dettagliata tutte le informazioni ricavate, in modo tale da costruire un flusso di lavoro coerente e corretto nel rispetto delle possibilità e delle funzionalità degli strumenti. Così facendo è nato il diagramma (e quindi il workflow) di ricerca OSINT oggetto di questa tesi.

2.2.1 LucidChart

Per la creazione del suddetto diagramma è stato utilizzato il software Lucidchart.[14] Questo software online si occupa della creazione di diagrammi, processi o strutture organizzative nell'ambito di un qualsiasi progetto di lavoro. Grazie ad un'ampia gamma di template e di protocolli preimpostati, tutti i suddetti processi vengono semplificati e velocizzati.

Sfruttando un linguaggio di comunicazione visivo, LucidChart crea un semplice ambiente collaborativo, nel quale tutti i membri di un team restano sempre allineati sugli sviluppi di un progetto.

Inoltre, essendo una piattaforma basata sul cloud, non sorgono problemi di compatibilità in quanto è possibile accedere al software indipendentemente dal sistema operativo (Windows, Mac e Linux), dal browser e dal dispositivo utilizzato.

Nella seguente figura viene mostrata una schermata di utilizzo di Lucidchart: al centro si trova il diagramma, disegnato sfruttando le impostazioni disponibili nella barra laterale; a sinistra sono invece riportate altre impostazioni riguardanti l'integrazione e la condivisione del diagramma con strumenti di terze parti.

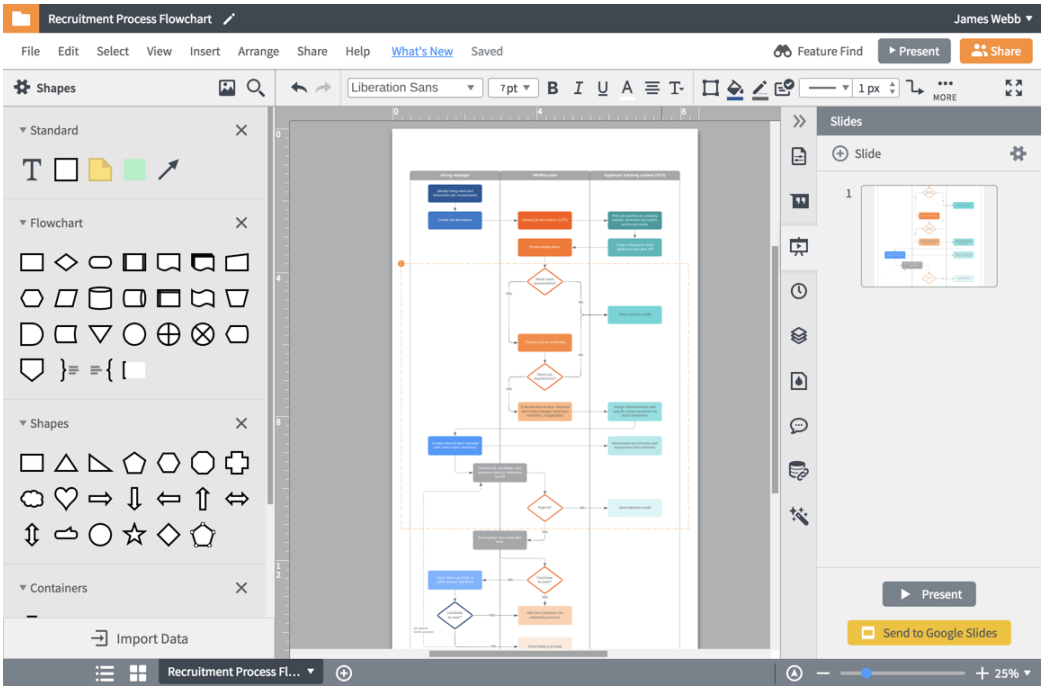


Figura 2.7: Schermata esempio per l'utilizzo di Lucidchart [14]

2.3 Ambiente di sviluppo e linguaggi utilizzati

Dopo l'analisi degli strumenti e dopo la realizzazione di un diagramma di ricerca OSINT completo e dettagliato volto a raccogliere il maggior numero di informazioni riguardanti uno specifico target, è possibile proseguire con la fase di implementazione del progetto.

2.3.1 Windows

Windows[62] è un sistema operativo proprietario, sviluppato dall'azienda Microsoft Corporation a partire dal 1985, caratterizzato da una semplice interfaccia grafica a desktop che consente la navigazione a "finestre".

Oltre alla gestione del computer e dei suoi programmi interni, Windows fornisce una serie di programmi gratuiti preinstallati tramite i quali si possono

svolgere facilmente operazioni di scrittura e lettura documenti, navigazione in rete, riproduzione di video, audio e brani musicali, stampa e condivisione di foto o documenti e tanto altro.

La seguente figura mostra l'evoluzione del sistema operativo Windows, partendo da Windows 1 fino ad arrivare all'ultima versione Windows 11.

Figura 2.8: Evoluzione del sistema operativo Windows: da Windows 1 fino a Windows 11 [15]

Nel corso degli anni, si possono considerare le versioni più significative e degne di nota le seguenti:

- *Windows 1 (1985)*: prima versione del sistema operativo, consente una semplice navigazione tra le finestre e le risorse del sistema.
- *Windows 95 (1995)*: viene integrato il DOS (Disk Operating System) e si inaugura la serie di sistemi operativi che potranno essere installati in ambienti operativi sia a 16 bit che a 32 bit.
- *Windows 8 (2012)*: le principali novità introdotte da questa versione coinvolgono la schermata home, l'interfaccia utente e la presenza di app preinstallate.

- *Windows 10 (2015)*: questa versione è la prima ad essere realmente multi piattaforma: dall'utilizzo del sistema operativo sul solo PC si passa all'inclusione di schermi touchscreen, console o notebook, adattando di conseguenza l'interfaccia utente.

- *Windows 11 (2021)*: ultima versione di Windows, la stessa utilizzata per lo sviluppo di questo progetto, nella quale il layout generale viene completamente rivoluzionato per permettere una più facile gestione delle finestre e delle applicazioni durante le esecuzioni simultanee.

2.3.2 GitLab

GitLab[16] è la piattaforma DevOps aperta, open-source e gestita da GitLab Inc, nata come singola applicazione che si occupa dell'intero ciclo di vita di sviluppo software, dalla creazione, allo sviluppo, al mantenimento e alla correzione di errori, per aiutare i team durante tutte queste fasi e aumentare la collaborazione, la visibilità, la qualità e la velocità del flusso di lavoro.

Con il termine DevOps [63] si intende una combinazione di "Development" (sviluppo) e "Operations" (operazioni), ossia una metodologia di lavoro che ha come obiettivo l'accelerazione del processo di creazione di un software puntando sulla comunicazione e collaborazione tra sviluppatori e addetti alle operations.

Seguendo questa filosofia di lavoro DevOps, nella seguente figura vengono riassunte quelle che sono le principali caratteristiche, funzionalità e potenzialità della piattaforma GitLab.

I vantaggi nell'utilizzo di un'unica piattaforma come GitLab per l'intero ciclo di vita di un software sono numerosi, tra cui:

- *Completezza e Continuità*: l'intero ciclo di vita di un software viene tenuto sotto controllo e gestito all'interno di un'unica piattaforma, utilizzando all'interno del sistema un unico set di strumenti, comuni tra i membri del team e tra le diversi fasi di sviluppo.

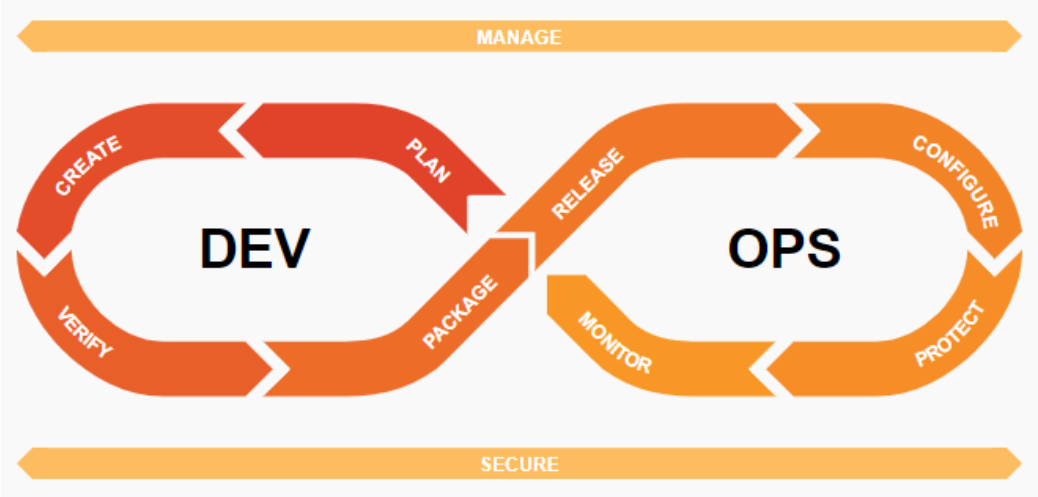


Figura 2.9: Schermata principali componenti IDE Spyder [16]

- *Trasparenza, Semplicità e Sicurezza*: ciascun membro del team deve imparare a gestire una sola piattaforma, tramite la quale effettuare le operazioni per lo sviluppo del software in maniera pulita, trasparente e tracciabile da ciascuno degli altri membri. Così facendo, vulnerabilità, violazioni di conformità o errori di programmazione saranno più facilmente rintracciabili.

Tutte le considerazioni fatte fino a questo momento non possono fare a meno che portare alla scelta di questo strumento come repository del progetto di tesi, per facilitare anche miglioramenti e/o integrazioni future.

2.3.3 Git per Windows

Considerando l'utilizzo del sistema operativo Windows 11, si pone ora il seguente problema: come fare per ottenere un dialogo semplice ed efficace tra il sistema operativo e il repository GitLab appena creato in modo da salvare mano a mano tutti i progressi del progetto?

A questo proposito, si può introdurre Git for Windows[17] ovvero uno strumento che riunisce tutte le funzionalità e potenzialità native del sistema Git SCM. Con quest'ultimo termine si intende un sistema di controllo delle

versioni, cioè un sistema che consente di eseguire diversi processi di sviluppo contemporaneamente sullo stesso progetto da parte di più utenti, tenendo traccia delle modifiche, dei cambiamenti e degli eventuali errori.

Dato che il sistema Git SCM è stato progettato per un ambiente Linux[64], il compito di Git for Windows non è altro quello di portare questa possibilità in un ambiente Windows, mettendo a disposizione degli utenti le seguenti funzionalità:

- **Git BASH**: questa funzionalità permette di emulare esattamente la shell BASH per eseguire Git da riga di comando come in un ambiente Linux. La seguente immagine mostra il processo di download ed inizializzazione di Git BASH, inserito all'interno di quella che sarà la riga di comando dalla quale partiranno le richieste. Nel caso specifico di questo progetto di tesi, la shell BASH emulata da Git è la versione 4.4.23.

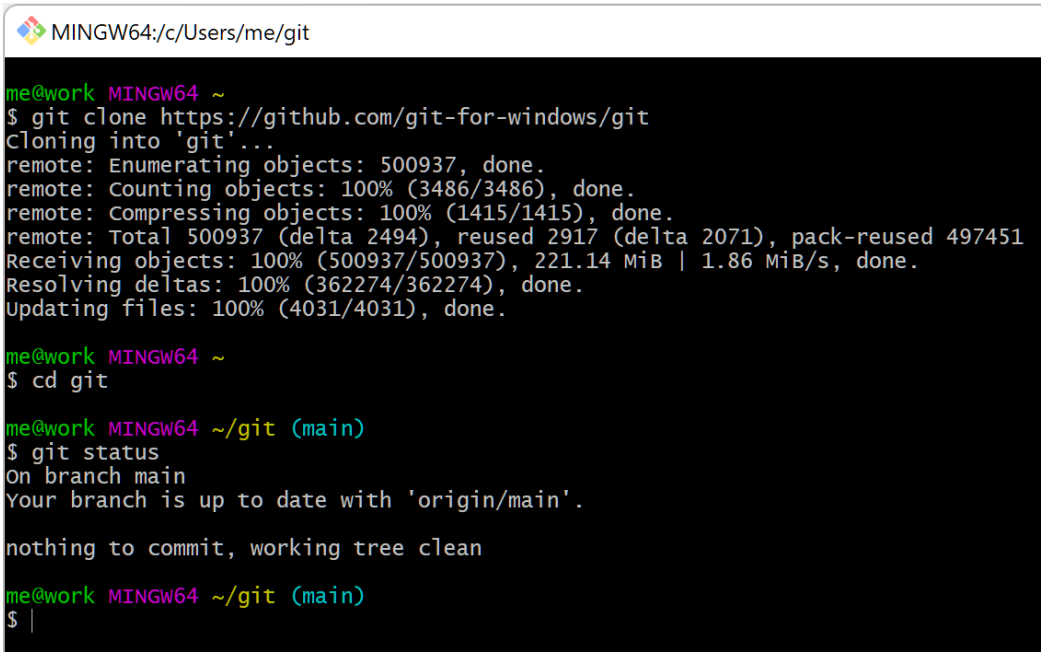


Figura 2.10: Schermata di utilizzo della Git BASH durante la sua prima installazione [17]

- **Git GUI**: questa funzionalità implementa un'interfaccia grafica fornita come alternativa alla standard Git BASH, in modo da poter comunque eseguire tutte le operazioni disponibili dalla riga di comando ma in maniera più facile ed intuitiva.

- **Integrazione della shell**: per facilitare l'integrazione e l'utilizzo di queste funzionalità, è possibile accedere a Git BASH o a Git GUI da una qualsiasi cartella Windows.

2.3.4 Spyder

Spyder[18] è il principale ambiente di sviluppo integrato (IDE) in campo scientifico, usufruibile in modalità completamente gratuita e open-source, scritto in Python e progettato per Python.

Questa piattaforma può essere installata su un qualsiasi sistema Windows, macOS e Linux: per i primi due sistemi è consigliabile utilizzare programmi di installazione "standalone", ovvero installatori autonomi che semplificano l'avvio e l'esecuzione evitando di effettuare tutti i processi manualmente; per Linux si consiglia di utilizzare la distribuzione Anaconda multi piattaforma, poiché oltre a Spyder vengono installati altri pacchetti solitamente utilizzati insieme. Se non si vuole procedere con l'installazione sul proprio dispositivo, è comunque possibile utilizzare Spyder visitando online la pagina Binder.

La seguente immagine mostra le principali componenti dell'IDE Spyder, descrivendo per ciascuno le proprie funzionalità e caratteristiche:

- **Editor**: la parte centrale di Spyder, nella quale viene scritto il codice Python. Si tratta di un editor multilingua, efficiente e di facile utilizzo e supporta operazioni come analisi e completamento automatico del codice, evidenziazione della sintassi, suddivisione orizzontale/verticale dell'ambiente di lavoro, suggerimenti per le chiamate e vai alla definizione e browser di funzioni/classi/metodi.

- **IPython Console**: la console IPython consente di eseguire comandi, righe e celle di codice o file, interagendo con i dati all'interno degli in-

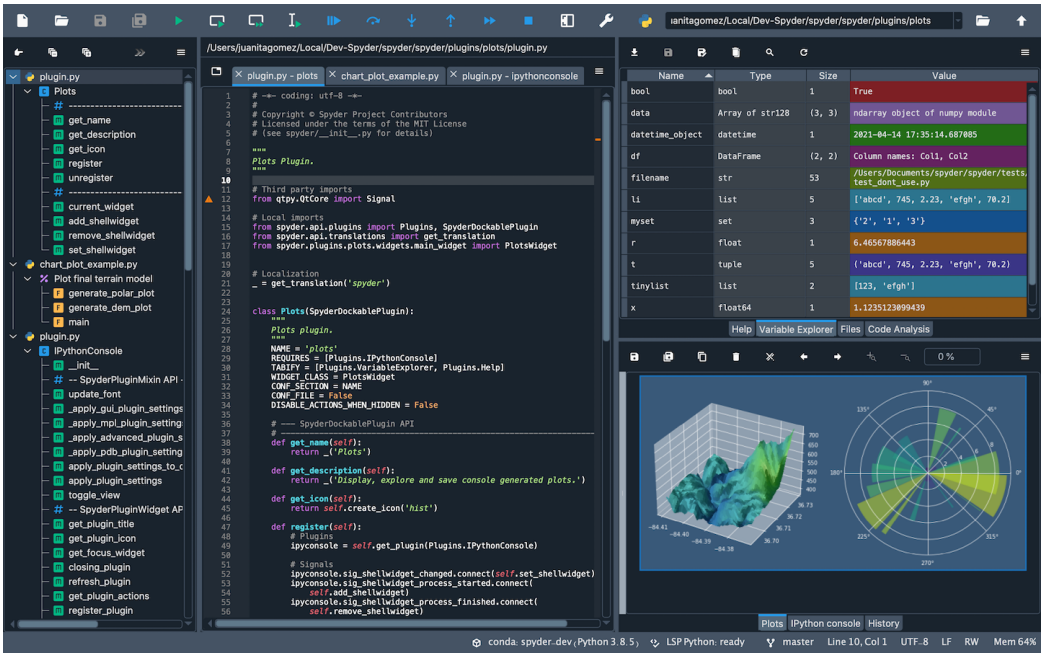


Figura 2.11: Schermata principali componenti IDE Spyder [18]

terpreti IPython. Inoltre, comunica e lavora in modo interattivo con le sezioni di debugging e plots. Si possono avviare più istanze di IPython: ciascuna verrà eseguita in un processo separato e sarà possibile effettuare operazioni senza creare interferenze con altre console in esecuzione. Inoltre, previa un'apposita procedura di configurazione, è possibile connettere alla console kernel esterni come ad esempio quelli gestiti da Jupyter Notebook.

- **Variable Explorer**: questo strumento permette di navigare, interagire e gestire in maniera veloce ed interattiva tutte le variabili e tutti gli oggetti presenti nel codice e contenuti nella sessione corrente della console IPython. A seconda della natura degli oggetti selezionati, verranno presentate nell'indice tutte le sue caratteristiche insieme ai valori corrispondenti, in modo tale da agevolare eventuali operazioni come modifica, cancellazione, duplicazione, inserimento e ridenominazione.

- **Plots**: all'interno di questa sezione è possibile interagire con tutte le

immagini, le figure e i grafici che vengono creati all'interno di Spyder. Ogni qualvolta che si cambia console o si esegue un nuovo codice, l'elenco delle figure visualizzate cambierà di conseguenza.

- **Debugger**: questa funzionalità permette di controllare il flusso di esecuzione del codice ed i punti di interruzione direttamente dalla IPython console, utilizzando l'apposita integrazione propria della console. Il debugger offre le funzionalità di evidenziazione della sintassi, completamento del codice, cronologia dei comandi ed è integrato nella sezione 'Breakpoints' che elenca il file, la riga e le condizioni di ogni punto di interruzione definito. Inoltre, è possibile accedere, modificare ed eliminare variabili locali, globali, grafici ed altri elementi in ogni breakpoint. Il punto esatto in cui si arresta il debugger viene indicato nell'editor con una freccia.

- **Help**: accedendo a questa sezione potrai visualizzare la documentazione di un qualsiasi oggetto che abbia una docstring direttamente all'interno di Spyder. Potrai ottenere le informazioni inserendo manualmente il nome dell'oggetto ricercato nella casella di testo apposita oppure direttamente all'interno dell'editor, grazie al pop up che apparirà passando sopra l'oggetto con il mouse.

2.3.5 Python

Python[65] è un linguaggio di programmazione di alto livello, open-source, rilasciato per la prima volta nel 1991 dal suo creatore Guido van Rossum ed attualmente gestito dall'organizzazione no-profit Python Software Foundation. Si tratta di un linguaggio multi piattaforma in quanto è in grado di supportare diversi paradigmi di programmazione come quella orientata agli oggetti, quella imperativa e quella funzionale.

Fornito di una ricca libreria di funzioni built-in, potenti costrutti per la gestione di operazioni ed eccezioni e considerando la gestione automatica della memoria, rientra tra i linguaggi più ricchi, facili e comodi da utilizzare.

Python è infatti progettato per essere facilmente intuibile sia da leggere che da implementare, utilizzando una sintassi pulita, snella, scorrevole unita a semplici e chiari costrutti.

Si tratta di un linguaggio interpretato perché può essere eseguito da qualsiasi piattaforma (macOS, Windows o Linux) previa installazione dello specifico interprete. L'interprete classico è scritto in C ed è chiamato CPython ma esistono anche altri interpreti che consentono l'integrazione con gli altri linguaggi.

Considerando le vaste funzionalità e potenzialità di questo linguaggio, lo si trova impiegato nello sviluppo di applicazioni e progetti di svariata natura. Infatti, grazie alle sue numerosissime librerie unite a tutti i package e/o ai moduli di terze parti, Python si trova ampiamente utilizzato nei seguenti ambiti:

- *Sviluppo Web*: nell'ambito dello sviluppo web, Python considera e gestisce sia la programmazione lato server (scrittura e fornitura di applicazioni web, siti o pagine web) sia la programmazione lato client (accesso ad applicazioni web, siti o pagine web). Esistono quindi numerosi strumenti e librerie apposite per il supporto, la realizzazione e la gestione di questi sistemi.

- *Accesso ai database*: tramite Python è possibile collegarsi praticamente a qualsiasi database esistente, a partire dai database relazioni, quelli non-relazionali fino al supporto per gli ODBC (Open Database Connectivity). La standard library di Python include l'interfaccia di accesso al database SQLite, ma installando gli appositi moduli è possibile gestire altri database come Oracle o MySQL.

- *Applicazioni desktop*: Python viene impiegato anche per lo sviluppo di interfacce grafiche o di applicazioni desktop, in quanto possiede numerosi framework per la creazione e la gestione di GUI che gli utenti andranno ad utilizzare.

- *Giochi e grafica 3D*: sempre più frequentemente si trova Python come linguaggio di sviluppo di giochi e piattaforme simili. In questo contesto, una particolare attenzione è da rivolgere a Pygame, ovvero un insieme di moduli open-source appositamente progettati per sviluppare videogiochi in maniera semplice e intuitiva.

- *Calcolo scientifico e numerico*: grazie a potenti ed efficienti librerie di calcolo, di elaborazione dati e di informazioni, Python viene ampiamente impiegato anche in ambito scientifico.

Queste caratteristiche, unite alla sua grande versatilità, hanno fatto di Python il linguaggio protagonista di questo elaborato di tesi, unito alla shell Bash che a seguito si andrà a descrivere ed analizzare. Ad ogni revisione del linguaggio si aggiungono nuove funzionalità che permettono così di tenere il passo con le nuove pratiche di sviluppo software. In questo elaborato di progetto, la versione di Python utilizzata è la 3.9[66].

2.3.6 Bash

Considerando Windows come sistema operativo di riferimento e dopo aver presentato lo strumento Git for Windows a suo supporto, si può introdurre il secondo strumento di programmazione utilizzato nella realizzazione di questo progetto di tesi.

BASH[67] (acronimo di Bourne Again SHell) è una shell testuale del sistema operativo GNU/Linux ideata da Stephen Bourne ed è l'evoluzione della shell standard /bin/sh di Unix Bash. La Bash è principalmente un interprete di comandi e come tale permette all'utente di comunicare con il sistema operativo attraverso funzioni predefinite oppure eseguendo comandi e/o script: i comandi vengono digitati direttamente tramite la linea di comando mentre gli script vengono realizzati utilizzando l'apposito linguaggio di scripting nativo.

Ma che cosa si intende con script? Gli script non sono altro che una serie di singoli comandi operativi riuniti a formare una sorta di programma

e contenenti anche variabili, strutture di controllo e altri elementi che caratterizzano i linguaggio di programmazione. A differenza di questi ultimi, gli script Bash non devono essere compilati per essere eseguiti poiché il software interprete del sistema che si occupa della loro esecuzione è in grado di comprenderli direttamente senza alcun passaggio intermedio.

Il compito principali della shell Bash è quello di eseguire tutti i comandi che le vengono passati tramite la linea di comando ma anche fungere da semplice linguaggio di programmazione che renda possibile svolgere operazioni più complesse e comandi combinati attraverso l'esecuzione degli script. In aggiunta, tra le maggiori funzionalità offerte dalla Bash si trova la ridirezione dell'input e dell'output, attraverso la quale è possibile eseguire più programmi in cascata passando come input dell'uno l'output dell'altro.

Proprio grazie a quest'ultima caratteristica, la shell Bash è stata scelta per essere utilizzata all'interno di questo elaborato di tesi, in modo tale da gestire al meglio gli input e gli output di ciascuno dei tools precedentemente descritti ed integrarli tra loro nel migliore dei modi.

Nonostante la Bash sia la shell più utilizzata in ambito Linux o Unix-Like, esistono altre shell più o meno specializzate nell'eseguire determinati programmi, gestire determinati ambienti di lavoro o nel migliorare funzionalità.

Capitolo 3

Analisi ed implementazione del progetto

All'interno di questo terzo ed ultimo capitolo si andrà dapprima ad analizzare il problema, illustrando il diagramma di lavoro creato, e successivamente si andrà a descrivere l'implementazione della soluzione trovata.

Nella prima parte verrà quindi mostrato il diagramma di lavoro creato sulla base degli strumenti analizzati in precedenza, spiegando tutte le scelte progettuali sulla base dei diversi input e output relativi a ciascuno strumento di ricerca.

Nella seconda parte, verrà invece descritta la soluzione a questo problema, ovvero verrà descritto il progetto di tesi sviluppato. Si inizierà quindi con una presentazione della sua struttura in termini generali per poi passare alla descrizione dei singoli dettagli implementativi che la costituiscono. Nel descrivere tutto ciò, oltre ai dettagli implementativi, verranno presentate anche schermate contenenti codici e risultati delle ricerche, per focalizzare al meglio il contributo apportato da ciascuno strumento all'interno del processo di ricerca.

Nella descrizione di alcuni passaggi e di alcune funzionalità verrà data per scontata la conoscenza delle tecnologie e degli strumenti impiegati, in quanto descritti ed analizzati nei capitolo precedenti.

3.1 Analisi del problema

Il progetto di tesi in questione si basa sullo sviluppo di un'integrazione tra strumenti appositamente ricercati e selezionati nell'ambito dell'Open Source INTelligence, con l'obiettivo di riportare all'utente quante più informazioni possibili circa un target specifico, avendo come input di ricerca un nome di dominio.

Per raggiungere al meglio questo obiettivo, dapprima si è eseguita un'operazione di comparazione tra gli strumenti selezionati, basandosi sugli input e sugli output relativi a ciascun strumento. Successivamente, sulla base di quanto appena studiato, è stato elaborato un diagramma di ricerca contenente un flusso di lavoro a cascata da seguire per l'implementazione del progetto.

La seguente immagine riporta il diagramma di lavoro realizzato per raggiungere il massimo grado di integrazione tra gli strumenti, nel rispetto della coerenza e correttezza delle informazioni che vi si possono ricavare.

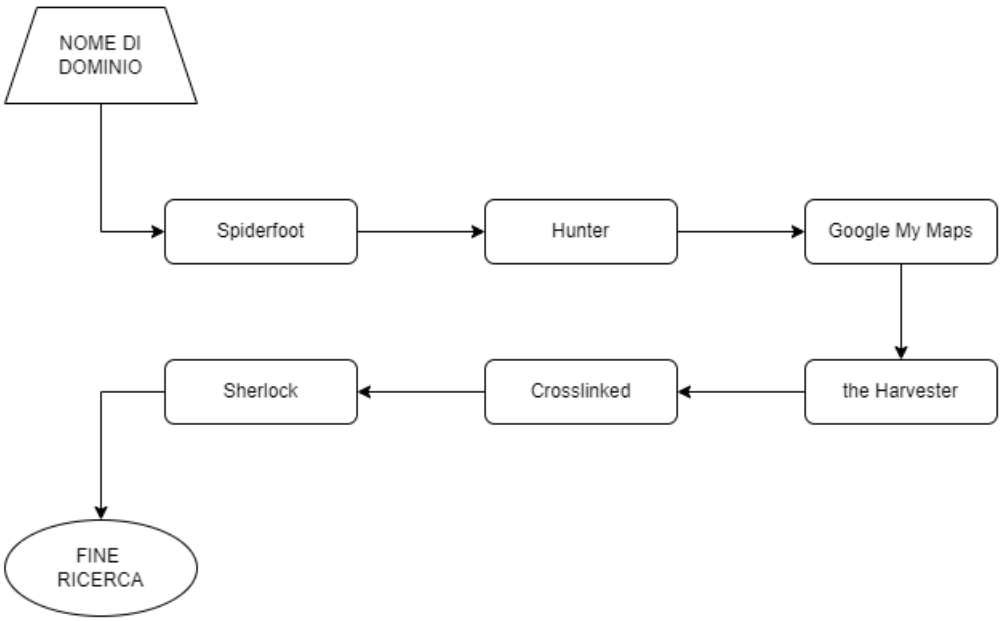


Figura 3.1: Diagramma di lavoro ed integrazione degli strumenti analizzati

Si andrà ora ad analizzare da più vicino questo diagramma, motivando ogni passaggio e la posizione degli strumenti al suo interno.

- *Spiderfoot*: la ricerca viene eseguita a partire dall'input inserito dall'utente ad inizio ricerca. Si è scelto di inserire Spiderfoot come primo strumento poiché grazie alle sue potenzialità e caratteristiche di ricerca, è in grado di fornire numerosi risultati della natura più disparata. Questi elementi costituiscono quindi un'ottima base di partenza per le analisi nei passaggi successivi.

- *Hunter*: viene effettuato un controllo circa la validità, l'attendibilità e la coerenza di alcune delle informazioni e delle risorse a propria disposizione, trovate tramite Spiderfoot nel passaggio precedente. Queste operazioni sono necessarie per effettuare una scrematura, dividendo ciò che potrebbe essere utile da ciò che non lo è, prima di passare alle operazioni di ricerca successive.

- *Google Maps*: a seconda della disponibilità di coordinate o di indirizzi geografici a questo punto della ricerca, è possibile eseguirne la conversione da un formato ad un altro, in modo tale da ricavare informazioni circa un determinato luogo.

- *the Harvester*: considerando l'input di partenza, vengono ricercate ulteriori informazioni ad esso associate come sottodomini, email, account e altre informazioni di rete. I dati trovati possono essere confrontati con quelli già in possesso ricavati dalle fasi precedenti oppure possono essere aggiunti come nuove informazioni.

- *Crosslinked*: vengono setacciati gli account LinkedIn per reperire email professionali o pubbliche associabili al nome di dominio fornito come input dall'utente o da eventuali altri domini noti. Crosslinked viene inserito proprio in questo punto perché considerando l'input aggiuntivo nome-cognome necessario per eseguire l'operazione di ricerca, potrebbe essere disponibile solo a questo punto della ricerca.

- *Sherlock*: partendo da un username come dato in input, vengono ricercati altri account nel web che vi corrispondano, sotto l'ipotesi che essi appartengano quindi allo stesso utente target. Questo strumento è inserito come ultimo tassello della ricerca perché, nel caso in cui venissero riscontrati nuovi account, è consigliabile ripetere le operazioni di ricerca precedenti per ciascuno di essi.

3.2 Implementazione del progetto

In questa sezione verrà presentata in maniera generale la struttura del progetto, la sua configurazione iniziale ed infine verrà descritta in maniera dettagliata l'implementazione dei singoli passaggi e dei singoli strumenti impiegati nella ricerca. Oltre ai dettagli implementativi, verranno riportate schermate contenenti i risultati forniti da ciascuno strumento al termine dalla propria esecuzione e frammenti di codice eseguibile.

3.2.1 Struttura generale e avvio della ricerca

Il progetto è composto da diversi file e moduli Python, ciascuno dei quali riporta l'implementazione di uno specifico strumento, e tutti vengono avviati a partire da un file principale: *automatizzazioneOSINT.py*.

All'apertura di questo file ha quindi inizio il processo di ricerca, come mostra la seguente immagine. Nelle prime righe di codice viene richiesto di inserire come input il target di riferimento, che non potrà più essere cambiato fino all'esecuzione di una nuova ricerca. Successivamente, vengono elencate tutte le possibili operazioni da eseguire, utilizzando una numerazione da 1 fino a 7: i primi sei numeri corrispondono all'esecuzione di uno specifico strumento mentre con l'ultimo numero si indica l'esecuzione del processo di ricerca completo, utilizzando quindi tutti gli strumenti in successione.

Infatti, per ottenere risultati ancora più completi ed affidabili, il progetto sviluppato prevede non solo di poter eseguire la ricerca in maniera unica e completa ma offre anche la possibilità di eseguire ciascuno degli strumenti sin-

golarmente, per tutti quei casi in cui è necessario concentrarsi maggiormente su un aspetto piuttosto che su un altro.

```
<<< RICERCA DI INFORMAZIONI RIGUARDANTI PERSONE O AZIENDE TRAMITE INDIRIZZO WEB >>>

Inserire il sito web di cui si vuole eseguire la scansione:
www.mariorossi.it

Scegliere la tipologia di ricerca da eseguire:

1) : ricerca utilizzando il tool SPIDERFOOT
Inserire nella nuova console un comando del tipo: ./spiderfoot_start_research.sh $TARGET

2) : ricerca utilizzando il tool HUNTER

3) : ricerca utilizzando il tool GOOGLE MY MAPS

4) : ricerca utilizzando il tool THEHARVESTER
Inserire nella nuova console un comando del tipo: ./theharvester_start_research.sh $TARGET

5) : ricerca utilizzando il tool CROSSLINKED
Inserire nella nuova console un comando del tipo: ./crosslinked_start_research.sh $TARGET $NOMERICERCA

6) : ricerca utilizzando il tool SHERLOCK
Inserire nella nuova console un comando del tipo: ./sherlock_start_research.sh $USERNAME

7) : ricerca utilizzando TUTTI i tool
```

Figura 3.2: Schermata iniziale per l'avvio della ricerca

Dopo aver brevemente descritto la struttura del progetto e dopo aver mostrato come appare la schermata di avvio della ricerca, si procederà ora con l'analisi dettagliata degli strumenti e delle scelte implementative.

3.2.2 Spiderfoot

A partire dal menu delle opzioni mostrato all'avvio, scegliendo il tasto '1', la ricerca verrà eseguita utilizzando unicamente Spiderfoot. Attraverso un primo script bash denominato *spiderfootstartresearch.sh* viene quindi richiesta l'apertura di una connessione al server, l'ascolto su una determinata porta e l'integrazione delle API da lui fornite, in modo tale da mantenere attiva la connessione per tutta la durata del processo di ricerca. Dopo aver quindi effettuato queste operazioni di configurazione iniziale, la ricerca può avviarsi.

Considerando il bacino dei moduli attivabili da Spiderfoot per rendere corpose e profonde le sue ricerche, non tutti risultano essere rilevanti per questo progetto di ricerca. Per questo motivo, all'interno di un secondo script *spi-*

derfootsettings.sh vengono impostati tutti i moduli che devono essere attivati durante il processo di ricerca, con la possibilità di essere modificati a seconda delle richieste e delle esigenze dell'utente.

Le seguenti due immagini mostrano gli script bash realizzati ed utilizzati per l'implementazione di quanto appena descritto: la prima fa riferimento al codice utilizzato nello script spiderfootstartresearch.sh mentre la seconda fa riferimento allo script spiderfootsettings.sh.

```
  GNU nano 4.3                                spiderfoot_start_research.sh
#!/bin/bash

#Script bash per esecuzione dei comandi e degli script di ricerca tramite il tool Spiderfoot

# FASE 1: accendo e metto in ascolto il server sulla porta 5001
python ./sf.py -l 127.0.0.1:5001 &

#FASE 2: eseguo le varie ricerche e salvo i risultati che ottengo su un file .txt
echo " "
echo "--- INIZIO DELLA RICERCA ---"
echo " "
./spiderfoot_settings.sh "$1" > spiderfoot_result.txt
echo " "
echo "--- ELABORAZIONE TERMINATA di \"$1\" ---"

#FASE 3: apro il file appena creato per visualizzare quanto appena trovato
echo " "
echo "--- VISUALIZZAZIONE DEL FILE CONTENENTE I RISULTATI DELLA RICERCA ---"
echo " "
start spiderfoot_result.txt

#FASE 4: chiudo la connessione al server eseguendo la kill del processo in background avviato in precedenza
kill $!
```

Figura 3.3: Script bash 'Spiderfootstartresearch.sh' per l'avvio della ricerca

```
  GNU nano 4.3                                spiderfoot_settings.sh
#!/bin/bash

#Script bash per esecuzione comandi Spiderfoot: vengono impostati i moduli interessati per la ricerca

python ./sf.py -m sfp_accounts,sfp_phone,sfp_email,sfp_names -s "$1" -o csv
```

Figura 3.4: Script bash 'Spiderfootsettings.sh' per le impostazioni della ricerca

Salvataggio e consultazione dei risultati ottenuti

Al termine della ricerca, i risultati ottenuti vengono mostrati sia sul terminale dal quale viene avviata la ricerca sia nel relativo file di testo. In questo ultimo modo, i risultati vengono visualizzati in un formato ben strutturato e manipolabile per facilitare eventuali altre operazioni su di essi.

Osservando la seguente immagine, si nota come le risorse trovate sono rese disponibili all'interno del file *spiderfootresult.txt* e vengono ordinatamente elencate secondo tre campi: 'Source', 'Type' ed infine 'Data'.

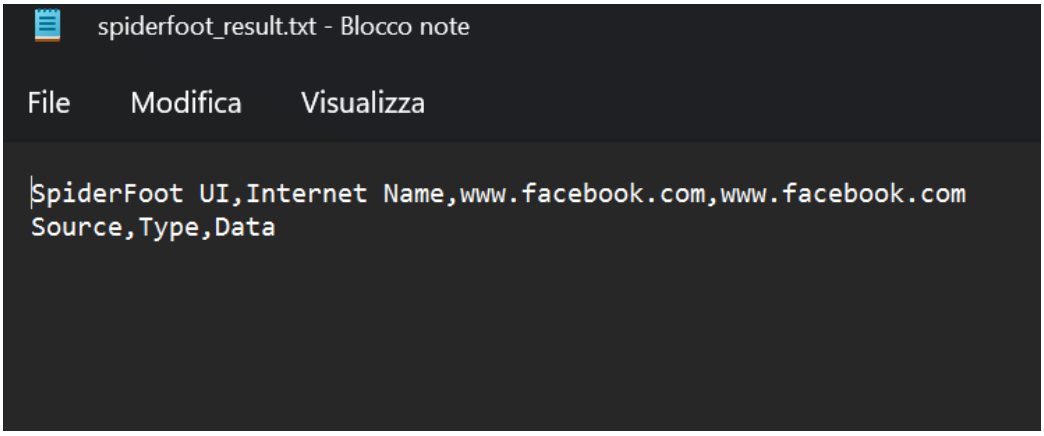


Figura 3.5: Risultati della ricerca tramite Spiderfoot

3.2.3 Hunter

Se dal menu principale viene invece selezionato il tasto '2' lo strumento selezionato per la ricerca è Hunter: da qui verrà aperta una seconda schermata, dalla quale sarà possibile scegliere quale operazione di controllo validità ed attendibilità eseguire in base alle informazioni a propria disposizione. Tra tutte le possibili opzioni, l'utente potrà quindi scegliere quelle più pertinente e in linea con i propri bisogni. Nell'implementazione di questo secondo passaggio di ricerca vengono eseguite svariate richieste HTTP e vengono utilizzate tutte le API messe a disposizione da Hunter, ciascuna specifica per l'operazione da eseguire. In questo contesto, viene richiesto l'inserimento di un'ulteriore chiave API personale, al fine di riconoscere ed associare i risultati ottenuti ad un determinato utente.

Nella seguente immagine viene mostrato il codice Python relativo all'implementazione di quanto appena detto. L'utente si trova davanti a tre possibili scelte, selezionabili tramite gli appositi comandi. Le singole operazioni

che si andranno ad elencare possono essere ripetute più volte e in maniera del tutto indipendente l'una dall'altra, in quanto è necessario il consenso esplicito da parte dell'utente nel voler terminare la sua ricerca o per passare ad una fase successiva della stessa.

```
print('Scegliere la tipologia di operazione da eseguire:')
print('1) : ricerca di indirizzi web dato il dominio in input')
print('2) : ricerca di indirizzi email professionali dato il dominio in input')
print('3) : verifica la validità di un indirizzo email')

scelta = input()

# 1. Domain search
if scelta == '1':
    r = requests.get("https://api.hunter.io/v2/domain-search?domain="+target+"&api_key="+api_key+"", stream=True)
    with open('./Tools/Hunter/domain_search_result.txt', 'wb') as fd:
        for chunk in r.iter_content(10):
            fd.write(chunk)
    with open('./Tools/Hunter/domain_search_result.txt', 'r+') as fd:
        content = fd.read()
        print("I risultati della ricerca sono i seguenti: \r\n\n" + content)
    fd.close()

# 2. Email finder
if scelta == '2':
    # variabili aggiuntive, necessarie per il coretto funzionamento di questa ricerca
    print('Inserire un nome di riferiment per il target: ')
    target_name = input()
    print('Inserire un cognome di riferimento per il target: ')
    target_surname = input()
    r = requests.get("https://api.hunter.io/v2/email-finder?domain="+target+"&first_name="+target_name+"&last_name="+target_surname+"&api_key="+api_key+"")
    with open('./Tools/Hunter/email_finder_result.txt', 'wb') as fd:
        for chunk in r.iter_content(10):
            fd.write(chunk)
    with open('./Tools/Hunter/email_finder_result.txt', 'r+') as fd:
        content = fd.read()
        print("I risultati della ricerca sono i seguenti: \r\n\n" + content)
    fd.close()

# 3. Email verifier
if scelta == '3':
    # variabile aggiuntiva, necessaria per il coretto funzionamento di questa ricerca
    print('Inserire un\'email di riferimento per il target: ')
    target_email = input()
    r = requests.get("https://api.hunter.io/v2/email-verifier?email="+target_email+"&api_key="+api_key+"")
    with open('./Tools/Hunter/email_verifier_result.txt', 'wb') as fd:
        for chunk in r.iter_content(10):
            fd.write(chunk)
    with open('./Tools/Hunter/email_verifier_result.txt', 'r+') as fd:
        content = fd.read()
        print("I risultati della ricerca sono i seguenti: \r\n\n" + content)
    fd.close()

print('\r\n --- Premere il tasto \'0\' per ripetere la ricerca; altro per procedere ---')
uscita = input()
if uscita == '0':
    hunter_research(target)
else:
    return
```

Figura 3.6: Implementazione ricerca tramite Hunter in linguaggio Python

- *Domain search*: viene effettuata una ricerca di indirizzi web avendo come input un nome di dominio. In questo caso non sono richiesti input secondari perché il nome di dominio viene passato come argomento.

- *Email Finder*: viene effettuata una ricerca di indirizzi email professionali o pubblici avendo come input un nome di dominio. Per l'esecuzione di questa operazione vengono richiesti due ulteriori paramenti in

input, ovvero nome e cognome della persona fisica che si presume essere associata al nome di dominio.

- *Email Verifier*: vengono effettuati tutti i controlli di validità di email inserendo come input un'email da verificare. Le email prese come input in questa terza opzione di ricerca potrebbero anche essere le stesse trovate nell'Email Finder.

Salvataggio e consultazione dei risultati ottenuti

A prescindere da quale sia la scelta presa dall'utente, tutti i risultati della ricerca vengono salvati all'interno dei relativi file di testo, ovvero il file *domainsearchresult.txt* per i risultati provenienti da Domain Search, il file *emailfinderresult.txt* per i dati provenienti da Email Finder ed infine il file *emailverifierresult.txt* per i risultati da email Verifier. Il formato scelto per il salvataggio è nuovamente un semplice file di testo per facilitare eventuali operazioni di manipolazioni ed integrazioni in altri passaggi o processi di ricerca.

3.2.4 Google My Maps

La ricerca tramite Google My Maps viene eseguita selezionando l'opzione '3'. Prima di procedere con l'avvio della ricerca, viene richiesto inserire la chiave API relativa al servizio di geocoding e di proprietà dell'utente stesso che ne richiede l'utilizzo, disponibile nel proprio account Google Developers[68], ovvero la piattaforma Google dedicata ed utilizzata dagli sviluppatori.

Una volta eseguita questa operazione preliminare, una seconda schermata mostra all'utente le due operazioni di conversione che possono essere avviate, a seconda dell'input a propria disposizione.

- *Coordinate GPS*: inserendo in ordine le coordinate GPS in gradi decimali rispettivamente latitudine e longitudine, utilizzando la chiave API fornita in precedenza, attraverso una specifica chiamata viene richiesto

di eseguire un'operazione di reverse geocoding, in modo da ottenere come output il corrispondente o i corrispondenti indirizzi fisici.

- *Indirizzo fisico*: inserendo come input un indirizzo fisico in un qualsiasi formato a discrezione dell'utente, sfruttando nuovamente la chiave API e la specifica chiamata, questa volta viene richiesto di eseguire un'operazione di geocoding, così da ottenere in output le coordinate GPS in formato decimale associate all'indirizzo inserito.

La seguente immagine riporta un frammento di codice Python riguardante le operazioni di conversione appena presentate e descritte: il tasto '1' permette di eseguire l'operazione di conversione da coordinate GPS a indirizzi fisici mentre il tasto '2' esegue la conversione inversa, ovvero da indirizzo fisico a coordinate GPS.

```
# coordinate GPS -> indirizzo fisico
if scelta == '1':
    print('Inserire le coordinate GPS (latitudine,longitudine)')
    lat = input("Latitudine: ")
    long = input("Longitudine: ")
    result = client_map.reverse_geocode([lat, long])

    #salvo i dati appena ottenuti
    file = open("./Tools/GoogleMaps/googlemaps.txt", "w")
    for i in result:
        file.write("%s\n" %i)
    file.close()

    #visualizzo i dati appena ottenuti
    with open('./Tools/GoogleMaps/googlemaps.txt', 'r+') as fd:
        content = fd.read()
        print("\r\nI risultati della ricerca sono i seguenti: \r\n\n" + content)
    fd.close()

# indirizzo fisico -> coordinate GPS
if scelta == '2':
    address = input("Inserire l\'indirizzo: ")
    result = client_map.geocode(address)

    #salvo i dati appena ottenuti
    file = open("./Tools/GoogleMaps/googlemaps.txt", "w")
    for i in result:
        file.write("%s\n" %i)
    file.close()

    #visualizzo i dati appena ottenuti
    with open('./Tools/GoogleMaps/googlemaps.txt', 'r+') as fd:
        content = fd.read()
        print("\r\nI risultati della ricerca sono i seguenti: \r\n\n" + content)
    fd.close()
```

Figura 3.7: Implementazione ricerca tramite Google My Maps in linguaggio Python

Queste operazioni possono essere eseguite in maniera dipendente o indipendente l'una dall'altra e possono essere ripetute più volte a discrezione

dell'utente e dei propri bisogni, in quanto è richiesto un esplicito consenso per poter terminare la ricerca o passare alla sua fase successiva.

Salvataggio e consultazione dei risultati ottenuti

I risultati delle ricerche tramite Google My Maps vengono salvati all'interno di un unico file di testo denominato *googlemaps.txt*. Per questo motivo, ogni volta che viene eseguita una nuova conversione, le informazioni trovate in precedenza devono essere immagazzinate da qualche parte per evitare di perdere i dati e tutto il contenuto informativo che ne deriva.

Le seguenti due immagini mostrano il contenuto del file 'googlemaps.txt' relativo alla ricerca avente come input rispettivamente le coordinate GPS e l'indirizzo fisico, in modo tale da provare il corretto funzionamento dello strumento di conversione. In questo caso, la ricerca avente come input l'indirizzo fisico riesce a riportare le coordinate geografiche di riferimento in maniera precisa mentre effettuando la ricerca inversa, il risultato è approssimativo in quanto non corrisponde all'indirizzo fisico precedente ma si trova nelle sue immediate vicinanze.


```
{'address_components': [{'long_name': '5', 'short_name': '5', 'types':
['street_number']}, {'long_name': 'Via G. Puntaroli', 'short_name': 'Via G. Puntaroli',
'types': ['route']}, {'long_name': 'Modigliana', 'short_name': 'Modigliana', 'types':
['locality', 'political']}, {'long_name': 'Modigliana', 'short_name': 'Modigliana',
'types': ['administrative_area_level_3', 'political']}, {'long_name': 'Provincia di
Forlì-Cesena', 'short_name': 'FC', 'types': ['administrative_area_level_2',
'political']}, {'long_name': 'Emilia-Romagna', 'short_name': 'Emilia-Romagna', 'types':
['administrative_area_level_1', 'political']}, {'long_name': 'Italy', 'short_name':
'IT', 'types': ['country', 'political']}, {'long_name': '47015', 'short_name': '47015',
'types': ['postal_code']}], 'formatted_address': 'Via G. Puntaroli, 5, 47015 Modigliana
FC, Italy', 'geometry': {'location': {'lat': 44.1569227, 'lng': 11.7906912},
'location_type': 'ROOFTOP', 'viewport': {'northeast': {'lat': 44.1582716802915, 'lng':
11.7920401802915}, 'southwest': {'lat': 44.1555737197085, 'lng': 11.7893422197085}}},
'place_id': 'ChIJwxPf6FBBKxMRhRFpW1oR5B0', 'plus_code': {'compound_code': '5Q4R+Q7
Modigliana, Province of Forlì-Cesena, Italy', 'global_code': '8FPH5Q4R+Q7'}, 'types':
['establishment', 'point_of_interest']}
```


Figura 3.8: Risultati della ricerca tramite Google My Maps con input coordinate GPS

{'address_components': [{'long_name': '13A', 'short_name': '13A', 'types':
['street_number']}, {'long_name': 'Via Sacchini', 'short_name': 'Via Sacchini', 'types':
['route']}, {'long_name': 'Modigliana', 'short_name': 'Modigliana', 'types': ['locality',
'political']}, {'long_name': 'Modigliana', 'short_name': 'Modigliana', 'types':
['administrative_area_level_3', 'political']}, {'long_name': 'Provincia di Forlì-Cesena',
'short_name': 'FC', 'types': ['administrative_area_level_2', 'political']}, {'long_name':
'Emilia-Romagna', 'short_name': 'Emilia-Romagna', 'types': ['administrative_area_level_
1', 'political']}, {'long_name': 'Italy', 'short_name': 'IT', 'types': ['country',
'political']}, {'long_name': '47015', 'short_name': '47015', 'types': ['postal_code']}],
'formatted_address': 'Via Sacchini, 13A, 47015 Modigliana FC, Italy', 'geometry':
{'location': {'lat': 44.1556077, 'lng': 11.7892603}, 'location_type': 'ROOFTOP',
'viewport': {'northeast': {'lat': 44.15695668029151, 'lng': 11.7906092802915},
'southwest': {'lat': 44.15425871970851, 'lng': 11.7879113197085}}}, 'place_id':
'ChIJNU0hgFpBKxMRqr2Ew-7tz90', 'plus_code': {'compound_code': '5Q4Q+6P Modigliana,
Province of Forlì-Cesena, Italy', 'global_code': '8FPH5Q4Q+6P'}, 'types':
['street_address']}

Figura 3.9: Risultati della ricerca tramite Google My Maps con input indirizzo fisico

3.2.5 the Harvester

theHarvester è il quarto strumento di ricerca e viene attivato utilizzando il pulsante '4'. La ricerca tramite questo strumento risulta essere molto semplice, veloce e di facile intuizione. Non vengono richieste infatti alcune chiavi API per l'avvio della ricerca e non è necessario aprire e stabilire in maniera diretta connessioni al server in quanto il tutto viene gestito in maniera autonoma ed indipendente da theHarvester stesso.

L'esecuzione della ricerca avviene quindi a partire dallo script bash *theharvesterstartresearch.sh*, nel quale sono stati precedentemente impostati dall'utente tutti i parametri di ricerca.

In questo caso viene richiesto di effettuare una ricerca in forma completa e generale, utilizzando quindi tutte le potenzialità e le funzionalità proprie di questo strumento. In alternativa, è sempre possibile modificare questa configurazione a discrezione dell'utente, seguendo le proprie esigenze e a seconda del dettaglio con cui si deve eseguire la ricerca.

La seguente immagine riporta lo script bash utilizzato per l'implementazione di questa ricerca: nelle prime righe si trova il codice che fa riferimento a quanto appena descritto mentre nelle ultime righe di codice viene descritta la procedura da seguire per il salvataggio dei risultati ottenuti.

```
  GNU nano 4.3                                  theharvester_start_research.sh
#! /bin/bash
#Script bash per esecuzione dei comandi e degli script di ricerca tramite il tool The Harvester

#FASE 1: eseguo le varie ricerche e salvo i risultati che ottengo nel file theharvester_result.html
echo " "
echo "--- INIZIO DELLA RICERCA ---"
echo " "
python theHarvester.py -d "$1" -b all -f theharvester_result
echo " "
echo "--- ELABORAZIONE TERMINATA di \"$1\" ---"

#FASE 2: apro il file appena creato per visualizzare quanto appena trovato
echo " "
echo "--- APERTURA VISUALIZZAZIONE DEI RISULTATI DELLA RICERCA SUL BROWSER ---"
start theharvester_result.html
```

Figura 3.10: Script bash 'theharvesterstartresearch.sh' per le impostazioni della ricerca

Salvataggio e consultazione dei risultati ottenuti

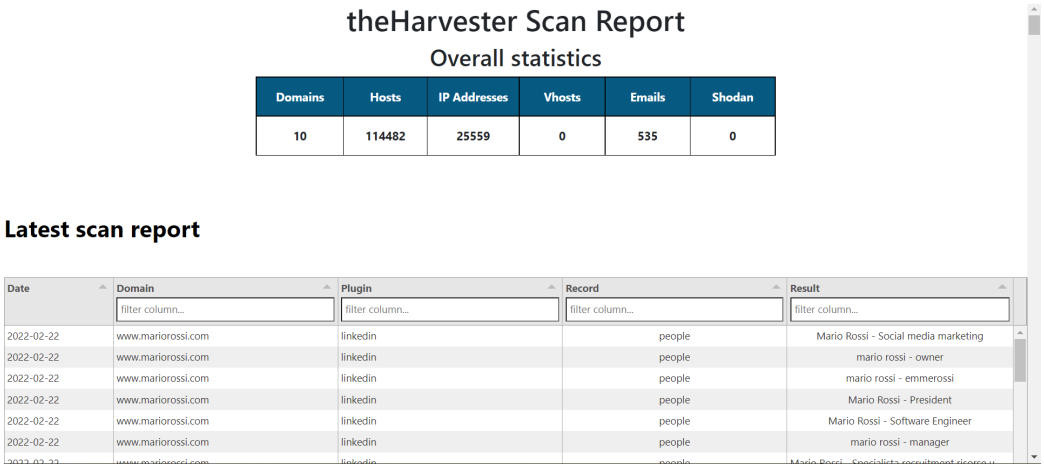
Al termine della ricerca, tutte le informazioni trovate vengono riportate sulla riga di comando in maniera abbastanza semplice e schematica ma, per consentire una migliore e completa navigazione tra i dati riportati dalla scansione, è possibile richiedere in maniera esplicita l'apertura del file *theharvesterresult.html* e visionare i risultati in un formato più accessibile.

All'interno di questo file, infatti, i risultati della scansione si trovano organizzati in una semplice ed ordinata struttura, pronti per essere consultati, manipolati o integrati all'interno di altre procedure di ricerca. Inoltre, trattandosi di un file con estensione .html, può essere facilmente accessibile e visualizzabile da un qualsiasi dispositivo e da un qualsiasi browser web. Si tratta quindi di un'ulteriore facilitazione ed incentivo per l'utilizzo di questa tipologia di visualizzazione dei risultati.

Le seguenti immagini riportano da più vicino il contenuto del file html in questione: come primo elemento si trova la tabella 'Overall Statistics' nella quale vengono mostrate le statistiche generali della scansione appena effettuata, riportando in termini quantitativi domini, host, indirizzi IP, email e tutti gli altri dati che ci si aspetta di ritrovare scorrendo ed analizzando l'intera pagina.

Successivamente si trova il corpo della ricerca: all'interno della tabella 'Latest Scan Report' vengono riportati tutti i dati della ricerca appena con-

clusa, suddivisi tra 'Date', 'Domain', 'Plugin', 'Record' e 'Result', in modo tale da avere per ciascuno di essi la data ed il dominio di inizio ricerca, lo strumento e la tipologia di dato analizzato ed infine il contenuto informativo. Al proprio interno, questa tabella possiede la funzionalità di filtro, per facilitare ulteriormente operazioni di ricerca e scrematura dei risultati.



theHarvester Scan Report

Overall statistics

Domains	Hosts	IP Addresses	Vhosts	Emails	Shodan
10	114482	25559	0	535	0

Latest scan report

Date	Domain	Plugin	Record	Result
	filter column...	filter column...	filter column...	filter column...
2022-02-22	www.mariorossi.com	linkedin	people	Mario Rossi - Social media marketing
2022-02-22	www.mariorossi.com	linkedin	people	mario rossi - owner
2022-02-22	www.mariorossi.com	linkedin	people	mario rossi - emmerossi
2022-02-22	www.mariorossi.com	linkedin	people	Mario Rossi - President
2022-02-22	www.mariorossi.com	linkedin	people	Mario Rossi - Software Engineer
2022-02-22	www.mariorossi.com	linkedin	people	mario rossi - manager

Figura 3.11: Tabelle dei risultati della ricerca tramite the Harvester

Scorrendo ulteriormente questa pagina, è possibile visualizzare altre tabelle come 'Previous Scan Report' e 'theHarvester Plugin Statistics': nella prima vengono riportate le statistiche generali riguardanti una precedente scansione mentre nella seconda si fa riferimento alle statistiche generali riguardanti tutte le ricerche effettuate tramite questo strumento. Inoltre, all'interno di ogni resoconto di scansione, vengono riportati data ed orario di generazione insieme al numero d'identificazione, in modo tale da poter confrontare a seguito quanto trovato senza creare confusione, errori o mescolanza di dati.

Come ultimo elemento di interesse all'interno di questa pagina si trovano due grafici a barre, riportati nella seguente immagine: nel primo grafico vengono riportate tutte le tipologie di dato raccolto insieme alla loro frequenza mentre nel secondo vengono riportate la velocità e le tempistiche con le quali questi dati vengono scoperti all'interno dal web. Per ciascuno di questi

grafici, viene messo a disposizione in alto a destra un menu di navigazione per poter effettuare eventuali operazioni di visualizzazione, manipolazione ed estrazione dei dati in maniera facile, veloce ed intuitiva.

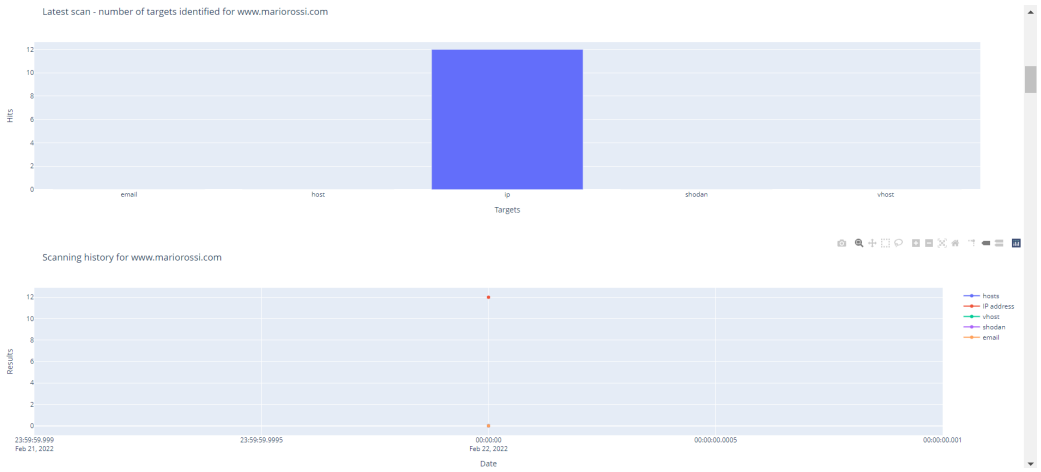


Figura 3.12: Grafici a barre dei risultati della ricerca tramite the Harvester

3.2.6 Crosslinked

Selezionando dal menu delle opzioni il numero '5' viene proposto come strumento di ricerca Crosslinked. Per l'avvio della ricerca, questo strumento fa uso dello script bash *crosslinkedstartresearch.sh* all'interno del quale vengono impostate tutte le configurazioni iniziali, sempre modificabili in un secondo momento a discrezione dell'utente e a seconda dei proprio obiettivi.

La seguente immagine mostra quindi la struttura della script Bash appena citato. Come prima azione viene richiesto di inserire tutti i parametri iniziali necessari all'avvio della ricerca: il nome di dominio viene ripreso dall'input fornito in precedenza mentre è necessario specificare un nome di riferimento da associare alla ricerca.

A questo punto, può avere inizio l'elaborazione della ricerca. In questa seconda fase vengono forniti due possibili formati di ricerca con i quali cercare di reperire gli account email associati al target:

'{first}.{last}@#DOMINIO_TARGET' '#NOME_RIFERIMENTO'

'#DOMINIO_TARGET\{f}{last}' '#NOME_RIFERIMENTO'

In tutti questi casi si può notare la presenza di 3 ulteriori paramenti richiesti, ovvero 'first', 'last' ed infine 'f': questi vengono utilizzati come segnaposto per la ricerca e fanno riferimento al possibile nome e cognome di un account email associato al dominio preso come input.

La ricerca viene ordinatamente eseguita in successione, partendo dall'elaborazione del primo formato di ricerca fino alla visualizzazione e al salvataggio dei risultati ottenuti nel relativi file di testo, concludendo poi con l'esecuzione delle medesime procedure anche per il secondo formato di ricerca.

```
  GNU nano 4.3                                  crosslinked_start_research.sh
#! /bin/bash
#Script bash per esecuzione dei comandi e degli script di ricerca tramite il tool Crosslinked

#FASE 1: eseguo le ricerche e salvo i risultati che ottengo nei due diversi file crosslinked_result1.txt e crosslinked_result2.txt
echo " "
echo "--- INIZIO DELLA RICERCA ---"
echo "Dominio target: \"$1\""
echo " "

#ricerca n.1
domain="$1"
python crosslinked.py -f {first}.{last}@$domain $2 -o crosslinked_result1.txt
#FASE 2: apro il file appena creato per visualizzare quanto appena trovato
echo " "
echo "--- APERTURA VISUALIZZAZIONE DEI RISULTATI DELLA RICERCA ---"
start crosslinked_result1.txt

#ricerca n.2
domain="$1"
python crosslinked.py -f $domain\\{f}{last} -t 45 -j 1 $2 -o crosslinked_result2.txt
#FASE 3: apro il file appena creato per visualizzare quanto appena trovato
echo " "
echo "--- APERTURA VISUALIZZAZIONE DEI RISULTATI DELLA RICERCA SUL BROWSER ---"
start crosslinked_result2.txt

#FASE 4: chiusura della ricerca
echo " "
echo "--- ELABORAZIONE TERMINATA di \"$2\" ---"
```

Figura 3.13: Script bash 'crosslinkedstartresearch.sh' per l'avvio della ricerca

Anche in questo caso non sono necessarie particolari accorgimenti e non vengono richieste chiavi API per l'esecuzione delle ricerche in quanto lo strumento Crosslinked gestisce il tutto in maniera autonoma ed indipendente.

Salvataggio e consultazione dei risultati ottenuti

Considerando l'impostazione delle due diverse strutture di ricerca, i risultati della ricerca vengono mostrati sia dalla riga di comando seguendo l'ordine di elaborazione ma anche in due appositi file di testo: *crosslinkedresult1* per

contenere i risultati della prima tipologia di ricerca e *crosslinkedresult2* per i risultati provenienti invece dalla seconda tipologia di ricerca, come è possibile vedere a partire dalle seguenti due immagini.

Questa prima immagine riporta il frammento di codice Python utilizzato per l'implementazione della funzionalità di salvataggio.

```python
#visualizzo i risultati della prima ricerca
with open('./Tools/Crosslinked/crosslinked_result1.txt', 'r+') as fd:
    content = fd.read()
    print("\r\nI risultati della prima ricerca del tipo '{first}.{last}@$domain' sono i seguenti: \r\n\n" + content)
fd.close()

#visualizzo i risultati della seconda ricerca
with open('./Tools/Crosslinked/crosslinked_result2.txt', 'r+') as fd:
    content = fd.read()
    print("\r\nI risultati della seconda ricerca del tipo 'domain\\{f}{last}' sono i seguenti: \r\n\n" + content)
fd.close()
```

Figura 3.14: Implementazione Python del salvataggio delle ricerche Crosslinked

Questa seconda immagine riporta invece le schermate dei due file di testo crosslinkedresult1.txt e crosslinkedresult2 appena presentati, ciascuno contenente i propri record di ricerca ordinatamente disposti per riga.

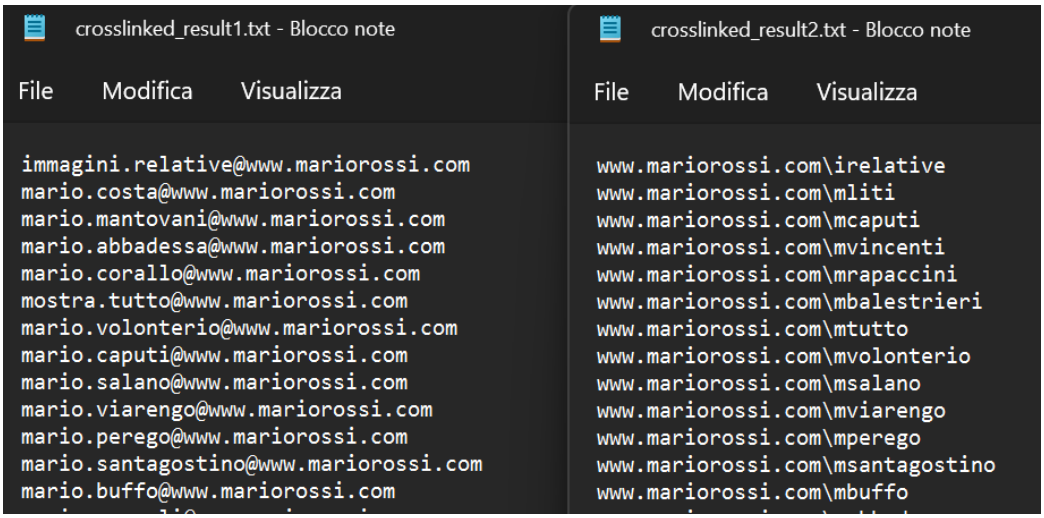


Figura 3.15: Risultati della ricerca in entrambi i formati tramite Crosslinked

3.2.7 Sherlock

Sherlock è l'ultimo strumento di ricerca Open Source INTelligence trattato in questo progetto di tesi e la sua ricerca viene avviata selezionando il pulsante '6'. Per il corretto avvio e la corretta esecuzione della ricerca è stato necessario implementare lo script bash *sherlockstartresult.sh*, all'interno del quale viene richiesto di eseguire l'elaborazione della ricerca non solo per un singolo ma per tutti i vari username che possono essere inseriti come input della ricerca.

```
  GNU nano 4.3                          sherlock_start_research.sh
#! /bin/bash
#Script bash per esecuzione dei comandi e degli script di ricerca tramite il tool Sherlock

#FASE 1: eseguo le ricerche e salvo i risultati che ottengo nel file sherlock_results
echo " "
echo "--- INIZIO DELLA RICERCA ---"
echo " "
python sherlock -fo sherlock_results $@

#FASE 2: apro il file appena creato per visualizzare quanto appena trovato
echo " "
echo "--- APERTURA VISUALIZZAZIONE DEI RISULTATI DELLA RICERCA ---"
start sherlock_results

#FASE 3: chiusura della ricerca
echo " "
echo "--- RICERCA TERMINATA ---"
```

Figura 3.16: Script bash 'sherlockstartresearch.sh' per l'avvio della ricerca

Salvataggio e consultazione dei risultati ottenuti

Per quanto riguarda il processo di salvataggio e consultazione dei risultati, lo stesso Sherlock è organizzato in maniera precisa ed ordinata.

A seconda del numero di username forniti come input ad inizio ricerca, vengono creati tanti file di testo, ciascuno denominato con l'username a cui fa riferimento per una maggiore chiarezza e fornire un riferimento immediato. All'interno di ciascun file, i dati sono disposti ad elenco, in maniera semplice ed ordinata, pronti ad essere eventualmente integrati in altri processi o in fasi di ricerca successive.

Tutti questi file sono contenuti all'interno della cartella *sherlockresult* e vengono mantenuti per tutta la durata desiderata dall'utente. Infatti, al-

l'avvio di una nuova ricerca, questi file non vengono eliminati o sovrascritti e rimangono quindi all'interno della stessa cartella. Per questo motivo, nel caso in cui fosse necessario eseguire più ricerche e suddividere i risultati, potrebbe essere opportuno creare ulteriori sottocartelle per separare i file di testo oppure spostarli in un'altra locazione a discrezione dell'utente.

La seguente immagine riporta i file contenuti all'interno della cartella sherlockresult, frutto di una ricerca avente come input 5 differenti username, ciascuno dei quali è riportato come nome del file di testo. Sulla destra è invece riportata la schermata contenente l'output della ricerca relativa al file *mariossi.txt*: ogni record è ordinatamente disposto ad elenco nella propria riga ed è formato dall'URL completo per poter accedere in maniera facile e diretta all'indirizzo web.

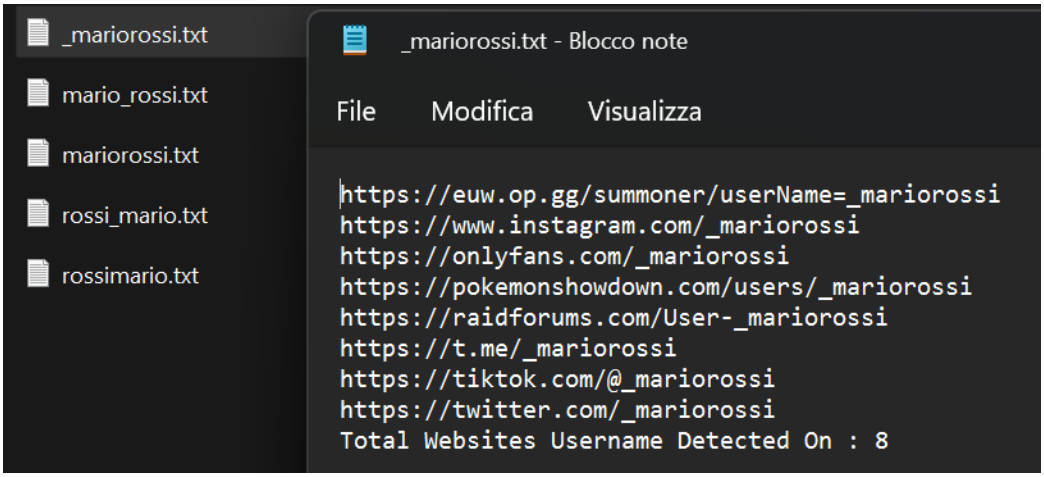


Figura 3.17: Risultati della ricerca tramite Sherlock con molteplici input

Al termine di tutte le ricerche, questa cartella viene quindi aperta in maniera automatica, lasciando all'utente la libertà di poter aprire i singoli file e di consultare i relativi risultati.

3.2.8 Ricerca completa

Come ultima operazione, inserendo come opzione il numero '7', è possibile effettuare una ricerca completa utilizzando nello stesso ordine tutti gli

strumenti appena analizzati.

Si ricorda che la struttura e l'ordine di questi strumenti non è casuale: ciascuno di essi è stato posizionato in un preciso momento della ricerca considerando le proprie potenzialità ed i propri argomenti di input e output, in modo da costituire un processo di ricerca ampio, ben strutturato e coerente, in grado di fornire risultati attendibili e facilmente consultabili nel rispetto dei relativi formati di output.

In questo ultimo caso non verrà quindi riportato nessun frammento di codice e nessuna immagine contenente i risultati delle ricerche, il tutto per evitare ridondanza, in quanto gli output degli strumenti sono stati analizzati nei paragrafi precedenti.

Conclusioni

In questi ultimi anni, la ricerca tramite tecnologie di Open Source INTelligence è cresciuta notevolmente sia da un punto di vista prestazionale che da un punto di vista dell'impiego. Tutte le realtà che ne facevano già ampio uso hanno visto grandi miglioramenti per quanto riguarda le funzionalità e le potenzialità di tutti gli strumenti e le tecnologie disponibili nel mercato ed utilizzabili. Allo stesso modo, la grande disparità di fonti rintracciabili unite alle nascita di nuove tecnologie, permettono alla ricerca OSINT di farsi strada ogni giorni in nuovi campi e nuovi ambienti di applicazione.

Per questo motivo, ci si aspetta di vedere un futuro sempre più brillante e luminoso in questo campo di ricerca e raccolta dati ed informazioni. All'interno di questo contesto, non ci si può però dimenticare che la ricerca OSINT è un processo molto complesso e che raggiungere un livello di automatizzazione precisa, corretta e completa di tutte le informazioni che si vorrebbero trovare non è facile da realizzare.

Risulta quindi di primaria importanza costruire solide mappe mentali e workflow ben strutturati, frutto di un profondo studio e di una profonda analisi, oltre a considerare sempre una minima presenza umana, necessaria ed indispensabile per la supervisione dei vari processi di ricerca e per l'analisi di alcune tipologie di dato non sempre di facile interpretazione.

Sulla base di queste prime considerazioni si fonda quindi il progetto di tesi sviluppato, mostrando come attraverso una laboriosa fase di analisi degli strumenti a propria disposizione si sia arrivati allo sviluppo di un'integrazione degli stessi, con l'obiettivo di raccogliere quante più informazioni circa

un target specifico in maniera fluida, coerente e corretta. Il tutto sempre nel rispetto dell'utente, al quale viene data la possibilità di modificare le operazioni di ricerca a seconda delle proprie esigenze ma anche la responsabilità circa la scelta dei risultati da prendere in considerazione per effettuare operazioni future.

Prendendo in esame il progetto svolto, risulta essere evidente la sua propensione all'integrazione di nuove e migliori funzionalità di ricerca, espandendo quindi lo spettro degli strumenti che possono essere utilizzati per l'esecuzione delle ricerche. In questo contesto, potrebbero quindi essere inseriti ulteriori strumenti sia di ricognizione che di controllo, posizionati in strategici punti all'interno del diagramma presentato nel capitolo precedente.

A seguito si elencano quelli che potrebbero essere gli strumenti di maggiore interesse per gli sviluppi futuri di questo progetto, aggiungendo anche quella che potrebbe essere la loro posizione ideale all'interno del flusso di lavoro:

- *Shodan*[69]*e Censys*[70]: questi motori di ricerca permettono di rintracciare tutti i dispositivi collegati alla rete Internet, appartenenti ad un determinato fornitore o geolocalizzati in un determinato punto; l'inserimento all'interno del flusso di ricerca si basa su una semplice e veloce integrazione, sfruttando sia i browser di ricerca ma anche le configurazioni dei singoli strumenti.

- *NameVine*[28]: controlla la reale presenza di un nome utente all'interno dei social media ma, più nello specifico, ricerca eventuali plausibili combinazioni dello stesso username in modo da aumentare notevolmente lo spettro delle ricerche; idealmente potrebbe essere posizionato prima dello strumento Sherlock, in quanto potrebbe fornire nuovi username su cui basare le ricerche.

- *Similarsites*[71]*e Similarweb*[47]: permettono di scoprire ulteriori siti web simili a quello fornito come input basandosi rispettivamente sul servizio fornito oppure sulla tipologia di contenuto presente all'interno

del sito; considerando la natura della sua ricerca e l'input richiesto, potrebbe essere inserito in un qualsiasi momento della ricerca.

Per ultimo, guardando il progetto nel suo insieme e focalizzando l'attenzione sull'aspetto prestazionale, non si può fare a meno di sottolineare come un miglioramento di tutte le tecniche di integrazione fino ad ora utilizzate per far dialogare tra loro i vari strumenti, possa sicuramente aumentare la velocità di ricerca e l'affidabilità dei risultati.

In modo specifico, per quanto riguarda questo ultimo aspetto di attendibilità dei risultati, bisogna ribadire come gli stessi strumenti impiegati ne stiano alla base: tutti gli strumenti inseriti all'interno di questo progetto vengono infatti gestiti in maniera autonoma ed indipendente dai propri produttori e contributori, i quali garantiscono ad ogni aggiornamento la migliore delle prestazioni in termini di risultati, efficienza ed attendibilità.

Per questo motivo, nello svolgimento di questo progetto di tesi, non sono stati effettuati test diretti mirati ad una valutazione prestazionale dello stesso, in quanto non sarebbe stata una procedura effettivamente efficace. Di maggiore importanza risulta quindi essere la scelta e la qualità degli strumenti che vengono integrati e la loro posizione all'interno del flusso di lavoro.

Pur non avendo seguito le procedure standard per la verifica dei risultati, le quali prevedono di avere a priori determinati dati in input ed i relativi dati in output in modo da verificare in maniera automatica la corrispondenza dei risultati, il progetto di tesi in questione non è stato lasciato alla casualità degli eventi e tutti i risultati ottenuti sono sempre stati controllati affinché rispettassero i criteri di correttezza, coerenza e attendibilità.

La corretta esecuzione di ciascuno degli strumenti e del progetto nella sua interezza è stata infatti controllata attraverso semplici test manuali, inserendo svariati dati in input e verificando che i dati in output ottenuti fossero coerenti con quelli che si sarebbe aspettato di ottenere.

A seguito di quanto appena detto, è stato sempre riscontrato un esito positivo al termine di ciascuna delle ricerche effettuate, in quanto gli strumenti sono sempre stati in grado di riportare risultati coerenti e consistenti.

Appendice

La seguente appendice si pone come obiettivo quello di semplificare al lettore la comprensione dell'intero elaborato di tesi e di facilitarne il suo utilizzo, mediante l'introduzione e la spiegazione passo dopo passo di un pratico esempio.

Riprendendo brevemente quanto detto nei capitoli precedenti, gli strumenti ed i passaggi chiave di questo sistema di ricerca OSINT, avente come obiettivo la ricerca del maggior numero di dati ed informazioni riguardanti un sito web preso come input, sono i seguenti:

- **Avvio del sistema**: viene richiesto di inserire il sito web target di riferimento per l'intera ricerca.
- **Indice delle ricerche**: viene mostrato l'elenco di tutte le possibili scelte di ricerca che possono essere eseguite, a partire dai singoli strumenti, ovvero Spiderfot, Hunter, Google My Maps, theHarvester, Crosslinked e Sherlock, fino alla ricerca completa che utilizza in maniera sequenziale tutti gli strumenti appena elencati.
- **Ricerca singola**: viene eseguita la ricerca unicamente con lo strumento selezionato.
- **Ricerca completa**: la ricerca viene svolta utilizzando uno dopo l'altro tutti gli strumenti a disposizione, seguendo l'ordine prestabilito nel diagramma di lavoro.

Per comprendere al meglio il lavoro svolto, si andranno ad analizzare quelli che sono i dati richiesti come input ed output per l'esecuzione della

ricerca. Nella spiegazione pratica dell'intero sistema, verrà preso come input di riferimento il sito web *www.mariorossi.com*. A seguito di questa prima scelta, verrà quindi richiesto di inserire il *numero '7'* all'interno dell'indice delle ricerche, in modo da selezionare e richiedere l'esecuzione della ricerca completa.

Il primo strumento con cui si avvierà la ricerca è *Spiderfoot*. Per l'esecuzione di questo strumento verrà aperta una nuova console, in modo tale da avere sempre disponibile come riferimento il flusso di lavoro da seguire nella console principale. Questo strumento richiede di inserire come input di ricerca il seguente comando: *./spiderfootstartresearch.sh www.mariorossi.com*. Al termine dell'esecuzione, l'output restituito consiste in dati di svariata natura come indirizzi email, username, nomi di persona o altri domini, il tutto visualizzato all'interno di un semplice file di testo.

Chiudendo in maniera esplicita questa seconda console, il sistema passa automaticamente alla ricerca tramite lo strumento successivo, ovvero *Hunter*. L'utente viene nuovamente sottoposto ad una scelta riguardante tre diverse tipologie di controllo, ciascuna avente un proprio input ed il relativo output:

- *ricerca indirizzi web*: come input viene richiesto di inserire un nome di dominio, il quale viene reperito automaticamente dal sistema utilizzando quello fornito ad inizio ricerca. In output vengono restituiti nuovi indirizzi web ad esso associati.

- *ricerca indirizzi email*: a partire dall'indirizzo web reperito in automatico dal sistema, viene richiesto l'inserimento di un nome e cognome; in questo caso si potrà quindi inserire *Mario* e *Rossi*. L'output restituito fa riferimento ai possibili indirizzi email professionali o pubblici associabili al target Mario Rossi e al nome di dominio www.mariorossi.com.

- *validità email*: come input viene richiesto di inserire un indirizzo email ad esempio *mariorossi@gmail.com* e vengono restituite in output svariate informazioni circa la sua validità, reperibilità e sicurezza nella rete.

Utilizzando un qualsiasi tasto al termine della ricerca è possibile procedere con l'esecuzione dello strumento di ricerca successivo, mentre la scelta del tasto '0' permette di ripetere nuovamente i controlli tramite Hunter.

Lo strumento successivo nel flusso di lavoro è *Google My Maps*. In questo caso, l'utente ha a propria disposizione due semplici operazioni di conversione:

- *da coordinate GPS a indirizzo fisico*: come input viene richiesto l'inserimento di coordinate GPS nel formato:*44.1558976, 11.7865661*, rispettivamente latitudine e longitudine. In output vengono restituiti tutti gli indirizzi fisici che fanno riferimento alle coordinate GPS fornite.

- *da indirizzo fisico a coordinate GPS*: come input viene richiesto l'inserimento di una stringa contenente un indirizzo fisico, per esempio: *Via Sacchini 13/a, 47015, Modigliana (FC), Italia*. In output vengono restituite le coordinate GPS in grado decimale associate all'indirizzo fornito.

Anche in questo caso, utilizzando un qualsiasi tasto al termine della ricerca, è possibile procedere con l'esecuzione dello strumento di successivo, mentre la scelta del tasto '0' permette di ripetere nuovamente le operazioni di conversione.

Il quarto strumento di ricerca è *theHarvester*. Per l'avvio di questa ricerca viene aperta una seconda console, all'interno della quale viene richiesto di inserire il seguente comando: *./theharvesterstartresearch.sh www.mariorossi.com*. Una volta terminata l'esecuzione della ricerca, l'output restituito viene ordinatamente mostrato all'interno di una pagina .html all'interno di un browser web. Le informazioni fornite sono le più disparate e riguardano indirizzi email, host, domini ed indirizzi IP associabili al target di riferimento.

Le operazioni di apertura e chiusura esplicita della console vengono richieste anche per effettuare le operazioni di ricerca tramite gli ultimi due strumenti, ovvero *Crosslinked* e *Sherlock*.

Dopo aver quindi chiuso la console generata dallo strumento theHarvetser, l'esecuzione della ricerca attraverso lo strumento Crosslinked prevede l'inserimento del seguente comando: *./crosslinkedstartresearch.sh www.mariorossi.com www.mariorossi.com*. L'output restituito da questa tipologia di ricerca consiste in semplici indirizzi email associati al target, i quali vengono suddivisi su due diversi file di testo, ciascuno contenete i propri elenchi di email differenziati per il formato di ricerca:

```
'{first}.{last}@mariorossi.com' 'www.mariorossi.com'
```

```
'mariorossi\{f}{last}' 'www.mariorossi.com'
```

Per ultimo, lo strumento Sherlock richiede l'inserimento dell'input seguendo il formato *sherlockstartresearch.sh $USERNAMES*, dove con $USERNAMES si fa riferimento a tutti i possibili nomi utente dei quali si vuole eseguire la ricerca. Per ciascuno di essi, verrà quindi creato uno specifico file di testo, nel quale verrà poi riportato per ciascun username l'elenco di tutti gli indirizzi URL ad esso associati trovati nel web.

Bibliografia

[1] Javier Pastor-Galindo, Pantaleone Nespoli, Félix Gómez Mármol, and Gregorio Martínez Pérez. The not yet exploited goldmine of osint: Opportunities, open challenges and future trends. *IEEE Access*, 8: 10282–10304, 2020. doi: 10.1109/ACCESS.2020.2965257.

[2] Fonti reperimento dati ed info osint, . URL `https://www.openpr.com/news/1904924/open-source-intelligence-market-know-factors-driving`.

[3] social media e social network osint, . URL `https://www.recuperomemorieelettroniche.com/www.recuperomemorieelettroniche.com/OSINT_2.html`.

[4] Justin Nordine. Osint framework, 2021. URL `https://osintframework.com/`.

[5] ilektrojohn. Geolocation osint tool. URL `https://www.geocreepy.com/`.

[6] Paterva. Maltego tool. URL `https://www.maltego.com/`.

[7] Caroline Kish and Tiffany Chiang. Osint integration tool. URL `https://www.osti.gov/biblio/1569427`.

[8] m8r0wn. Crosslinked: il miglior strumento di enumerazione di linkedin di cui non hai mai sentito par-

lare, 2021. URL https://infosecwriteups.com/crosslinked-the-greatest-linkedin-enumeration-tool-youve-never-heard-of-86a90

[9] URL https://www.google.it/maps/preview.

[10] URL https://hunter.io/.

[11] Haki. Sherlock – osint per hacker, 2020. URL https://deepadvice.it/osint/sherlook-tool-osint-per-hacker/.

[12] Cyberpunk. Osint collection and reconnaissance tool – spiderfoot, 2020. URL https://www.cyberpunk.rs/osint-collection-and-reconnaissance-tool-spiderfoot.

[13] Ethical Tools. theharvester best osint tool, 2020. URL https://ethicaltools.gitbook.io/subdomainfinder/theharvester-best-osint-tool.

[14] Lucidchart. Lucidchart. URL https://www.lucidchart.com/pages/it/prodotto?gclid=Cj0KCQiArt6PBhCoARIsAMF5waiL9XRH67sWKznAG0tQTuDspLaYGgHHYObQ8KF-DSJFIjVKOC9GtuwcB&km_CPC_AdGroupID=99270039979&km_CPC_AdPosition=&km_CPC_CampaignId=9594860760&km_CPC_Country=20553&km_CPC_Creative=424699413281&km_CPC_Device=c&km_CPC_ExtensionID=&km_CPC_Keyword=lucidchart&km_CPC_MatchType=e&km_CPC_Network=g&km_CPC_TargetID=aud-833150265254%3Akwd-33511936169&km_CPC_placement=&km_CPC_target=&utm_campaign=_chart_it_allcountries_mixed_search_brand_exact_&utm_medium=cpc&utm_source=google.

[15] Windows logo ed evoluzione, . URL https://loghi-famosi.com/windows-logo/.

[16] Gitlab, . URL https://about.gitlab.com/.

[17] Git per windows, . URL https://gitforwindows.org/.

[18] Spyder. Spyder. URL https://www.spyder-ide.org/.

[19] Ferruccio Diozzi. *Glossario di biblioteconomia e scienza dell'informazione*. Editrice Bibliografica, 2021. ISBN 9788893573177.

[20] Giovanni Nacci. *Open source intelligence abstraction layer: Proposta per una Teoria generale dell'intelligence delle fonti aperte*. Edizioni Epoké, 2014. ISBN 9788893573177.

[21] Significato della metodologia di ricerca (cos'è, concetto e definizione), 2022. URL https://it.encyclopedia-titanica.com/significado-de-metodolog-de-la-investigaci-n#menu-1.

[22] David Bisson. 10 open-source intelligence tools (that actually work with your existing security software), 2021. URL https://securityintelligence.com/articles/10-open-source-intelligence-tools-existing-security-software/.

[23] URL https://haveibeenpwned.com/.

[24] URL https://pipl.com/.

[25] URL https://knowem.com/.

[26] . URL https://www.namecheckr.com/.

[27] URL https://usersearch.org/index.php.

[28] . URL https://namevine.com/.

[29] URL https://www.lullar.com/.

[30] URL https://thatsthem.com/.

[31] URL https://www.spokeo.com/.

[32] URL http://www.yasni.com/.

[33] URL https://www.geni.com/.

[34] URL https://www.familysearch.org/tree/overview.

[35] URL http://wikimapia.org/.

[36] URL https://www.bing.com/maps.

[37] URL https://www.gps-coordinates.net/.

[38] URL https://eos.com/landviewer/.

[39] URL https://www.terraserver.com/.

[40] URL https://www.iplocation.net/.

[41] URL https://viewdns.info/.

[42] URL https://iknowwhatyoudownload.com/.

[43] URL https://securitytrails.com/dns-trails.

[44] . URL https://whoisology.com/.

[45] URL http://web.archive.org/web/.

[46] . URL https://who.is/.

[47] URL https://www.similarweb.com/.

[48] URL https://spyse.com/tools/subdomain-finder.

[49] Carol Verde. Che cos'è la social media intelligence e dove può condurre un business, 2019. URL https://www.ninjamarketing.it/2019/10/15/che-cose-la-social-media-intelligence/.

[50] Mirko Lapi. Open source intelligence: a cosa serve e come può essere impiegata per il sociale. 2021. doi: https://www.agendadigitale.eu/sicurezza/open-source-intelligence/.

[51] Laura Zanotti. Cosa sono le api e quale impatto hanno sul business, 2021. URL https://www.zerounoweb.it/software/erp-crm-scm/cosa-sono-le-api-e-quale-impatto-hanno-sul-business/.

[52] Forella Belcore. Quali le sanzioni previste per chi opera sprovvisto della licenza ex art. 134 tulps?, 2021. URL https://www.forensicnews.it/quali-le-sanzioni-previste-per-chi-opera-sprovvisto-della-licenza-ex-art-134-

[53] TULPS, . URL https://www.brocardi.it/testo-unico-pubblica-sicurezza/titolo-iv/art134.html.

[54] TULPS, . URL https://www.brocardi.it/testo-unico-pubblica-sicurezza/titolo-iv/art140.html.

[55] Paolo Palmieri. L'osint come strumento a disposizione del difensore. 2021. doi: https://www.cyberlaws.it/en/2021/osint-strumento-difensore/.

[56] Crosslinked tool. URL https://github.com/m8r0wn/CrossLinked.

[57] Stefano Regazzi. Google maps: cos'è, come funziona, come si usa e tutto quello che bisogna sapere, 2021. URL https://techprincess.it/google-maps-come-funziona-guida-completa/.

[58] Google Maps. Documentazione: geocoding. URL https://developers.google.com/maps/documentation/geocoding/overview.

[59] Sherlock tool. URL https://github.com/sherlock-project/sherlock.

[60] Steve Micallef. Spiderfoot tool. URL https://www.spiderfoot.net/.

[61] The harvester tool. URL https://github.com/laramies/theHarvester.

[62] Microsoft. Windows, . URL https://www.microsoft.com/it-it/windows/.

[63] RedHat. Cos'è devops, 2018. URL https://www.redhat.com/it/topics/devops.

[64] Linuxiano. Cos'è git, 2017. URL https://linuxiano.altervista.org/2017/12/git-scm/.

[65] Python. URL https://www.python.it/.

[66] Python. URL https://docs.python.org/3.9/.

[67] Andrea Minini. Cos'è la bash. URL https://www.andreaminini.com/linux/bash.

[68] Google. Google developers. URL https://developers.google.com/.

[69] Shodan. Shodan. URL https://www.shodan.io/.

[70] Censys. Censys. URL https://censys.io/.

[71] Similarsites. Similarsites. URL https://it.similarsites.com/.

[72] Michelangelo Di Stefano. Ricerca investigativa: le metodologie osint/socmint sulle fonti aperte. URL https://www.altalex.com/documents/news/2020/02/14/ricerca-investigativa-metodologie-osint-socmint-sulle-fonti-aperte.

[73] Hector Pellet, Stavros Shiaeles, and Stavros Stavrou. Localising social network users and profiling their movement. *Computers Security*, 11 2018. doi: 10.1016/j.cose.2018.10.009.

[74] Michael Glassman and Min Ju Kang. Intelligence in the internet age: The emergence and evolution of open source intelligence (osint). *Computers in Human Behavior*, 28(2):673–682, 2012. ISSN 0747-5632. doi: https://doi.org/10.1016/j.chb.2011.11.014. URL https://www.sciencedirect.com/science/article/pii/S0747563211002585.

[75] Stefano Scaini Elisa Riservato. Osint – intelligence da fonti aperte: un'incessante evoluzione dalle origini ai giorni nostri. 2018. doi: https://www.safetysecuritymagazine.com/articoli/osint-intelligence-da-fonti-aperte-unincessante-evoluzione-dalle-origini-ai-giorni-nostri/.

[76] Francesco Bechis. L'intelligence solidale? perché nasce osintitalia, 2021. URL `https://formiche.net/2021/04/lintelligence-solidale-perche-nasce-osintitalia/`.

[77] Microsoft. Windows 7, . URL `https://www.microsoft.com/it-it/windows/windows-7-end-of-life-support-information`.

[78] . URL `https://www.okpedia.it/windows`.

Ringraziamenti

Prima di tutto ringrazio la mia famiglia e le persone a me più care, che mi hanno sempre sostenuta nonostante i miei sbalzi di umore e mi hanno sempre spronata a dare il massimo, anche nei momenti più bui e difficili.

Ringrazio poi tutti gli amici più sinceri che mi sono stati vicino e che hanno creduto in me in questo percorso, spendendo qualche parole e parte del loro tempo nei momenti di sconforto e condividendo i piccoli e grandi successi.

Grazie anche ai miei compagni di università per aver alleggerito ogni lezione con tante risate, per gli appunti passati e per aver condiviso crisi di ogni genere in vista degli esami.

Infine vorrei ringraziare il Prof. Vittorio Ghini, nel ruolo di relatore di tesi, per la sua prontezza, la sua disponibilità e tutto il suo aiuto durante questo elaborato di tesi e ancora Marco Canducci, nel ruolo di correlatore di tesi, per l'impegno, il sostegno e per tutti i suoi preziosi consigli sia durante il percorso di tirocinio che durante questo percorso di tesi.