

Indice generale

Introduzione	xvii
Perché ho scritto questo libro.....	xvii
Che cosa scoprirete.....	xviii
Parte I – Fonti e dataset	xviii
Parte II – Gli strumenti del mestiere	xix
Parte III – Programmazione Python.....	xx
Parte IV – Dati strutturati	xx
Parte V – Casi di studio.....	xxi
Appendici.....	xxi
Di che cosa avrete bisogno	xxi
L'autore	xxii
La revisione tecnica.....	xxii
Ringraziamenti	xxiii
Parte I Fonti e dataset.....	1
Capitolo 1 Protezione delle fonti e di voi stessi.....	3
Comunicazioni in sicurezza con le fonti.....	4
Lavorare con dati pubblici.....	4
Proteggere le informazioni riservate.....	5
Ridurre al minimo le tracce digitali	5
Lavorare con hacker e informatori	6
Archiviazione sicura dei dataset	7
Dataset a bassa riservatezza	7
Dataset a media riservatezza	8
Dataset ad alta riservatezza	9
Autenticazione dei dataset.....	10
Il dataset AFLDS.....	11
La chat del gruppo Twitter di WikiLeaks	12
Oscuramento	14

Quali dati pubblicare	15
Che cosa oscurare	15
Effettuare richieste di commento	17
Gestori di password.....	18
Crittografia del disco	21
Esercizio 1.1 – Crittografare il disco interno	22
Windows.....	22
Mac OS.....	23
Linux	24
Esercizio 1.2 – Crittografare un disco USB	25
Windows.....	25
Mac OS.....	28
Linux	28
Protezione dai documenti dannosi	28
Esercizio 1.3 – Installare e utilizzare Dangerzone.....	30
Riepilogo	32

Capitolo 2 Acquisizione dei dataset33

La fine di WikiLeaks	34
Distributed Denial of Secrets.....	35
Download di dataset con BitTorrent.....	36
Le origini del dataset BlueLeaks	38
Esercizio 2.1 – Download del dataset BlueLeaks.....	38
Comunicazioni con app di messaggistica crittografate.....	39
Esercizio 2.2 – Installazione ed esercitazione sull’uso di Signal.....	41
Crittografia dei messaggi con PGP.....	41
Mantenimento dell’anonimato online con Tor e OnionShare	42
Esercizio 2.3 – Provare Tor e OnionShare.....	45
Comunicazioni con la mia fonte Tea Party Patriots	46
Altre opzioni per acquisire dataset dalle fonti	47
Unità USB crittografate.....	47
Server privati virtuali	48
Sistemi di segnalazione per informatori	49
Riepilogo	50

Parte II Strumenti del mestiere.....51

Capitolo 3 L’interfaccia a riga di comando53

Che cos’è la riga di comando	54
La shell.....	54
Utenti e percorsi.....	54
Privilegi degli utenti	56
Esercizio 3.1 – Installare Ubuntu in Windows	57
Utilizzo di base della riga di comando.....	59

Apertura del terminale.....	59
Cancellazione della schermata e uscita dalla shell.....	60
Esplorazione di file e directory.....	61
Navigazione su percorsi relativi e assoluti.....	63
Cambio di directory.....	63
Utilizzo dell'argomento help.....	64
Accesso alle pagine man.....	65
Suggerimenti per la navigazione nel terminale.....	65
Immissione di comandi con completamento	
tramite tabulazione.....	65
Editing dei comandi.....	66
Uso degli spazi nei nomi di file.....	67
Utilizzo delle virgolette singole per racchiudere le doppie	
virgolette.....	67
Installazione e disinstallazione del software	
con i gestori di pacchetti.....	68
Esercizio 3.2 – Gestione dei pacchetti con Homebrew	
su macOS.....	69
Esercizio 3.3 – Gestione dei pacchetti con apt su Windows	
o Linux.....	71
Esercizio 3.4 – Esercitarsi nell'uso della riga di comando	
con cURL.....	74
Scaricare una pagina web con cURL.....	74
Salvare una pagina web in un file.....	75
File di testo e file binari.....	75
Esercizio 3.5 – Installare l'editor di testo VS Code.....	76
Esercizio 3.6 – Scrivere il primo script della shell.....	77
Fate accesso al vostro disco USB.....	78
Create una cartella per gli esercizi.....	78
Aprirete un workspace VS Code.....	79
Scrivete lo script della shell.....	80
Eseguite lo script della shell.....	80
Esercizio 3.7 – Clonare il repository GitHub del libro.....	81
Riepilogo.....	82

Capitolo 4 Esplorazione dei dataset nel terminale83

I cicli.....	83
Esercizio 4.1 – Decomprimere il dataset BlueLeaks.....	85
Decomprimere i file su macOS o Linux.....	85
Decomprimere i file su Windows.....	88
Organizzare i file.....	89
Come l'hacker ha ottenuto i dati BlueLeaks.....	90
Esercizio 4.2 – Esplorare BlueLeaks dalla riga di comando.....	92
Calcolare la quantità di spazio su disco utilizzata dalle cartelle...92	

Usare le pipe e ordinare l'output.....	93
Creare un elenco dei nomi di file contenuti in un dataset.....	95
Contare i file contenuti in un dataset	95
Esercizio 4.3 – Trovare rivelazioni in BlueLeaks con grep	96
Filtrare i documenti che menzionano il termine antifa	96
Filtrare determinati tipi di file	97
Usare grep con le espressioni regolari.....	98
Cercare file in blocco con grep	98
Dati crittografati nel dataset BlueLeaks.....	100
Analisi dei dati con server nel cloud	102
Esercizio 4.4 – Configurare un VPS.....	104
Generare una chiave SSH	104
Aggiungere la vostra chiave pubblica al provider cloud	105
Creare un VPS	106
SSH nel vostro server.....	106
Iniziare una sessione Byobu.....	107
Installare gli aggiornamenti	108
Esercizio 4.5 – Esplorare il dataset Oath Keepers da remoto	109
Riepilogo	113

Capitolo 5 Docker, Aleph e le ricerche nei dataset115

Docker e i container Linux	116
Esercizio 5.1 – Inizializzare Docker Desktop su Windows e macOS.....	117
Esercizio 5.2 – Inizializzare il motore Docker su Linux.....	118
Esecuzione di container con Docker	119
Esecuzione di un container Ubuntu.....	119
Come elencare ed eliminare i container	120
Montaggio e rimozione di volumi	121
Passaggio di variabili di ambiente	124
Esecuzione del software per server	125
Come liberare spazio su disco	127
Esercizio 5.3 – Eseguire un sito WordPress con Docker Compose	127
Creare un file docker-compose.yaml	128
Avviare il sito WordPress	129
Introduzione ad Aleph	130
Esercizio 5.4 – Eseguire Aleph localmente nei container Linux....	131
Utilizzo dell'interfaccia web e a riga di comando di Aleph.....	133
Indicizzazione dei dati in Aleph.....	135
Esercizio 5.5 – Indicizzare una cartella di BlueLeaks in Aleph.....	135
Montare i dataset nella shell Aleph	136
Indicizzare la cartella icefishx	136

Controllare lo stato dell'indicizzazione	137
Eplorazione di BlueLeaks con Aleph	139
Funzionalità aggiuntive di Aleph	140
Server Aleph dedicati	142
Riepilogo	143
Capitolo 6 Leggere i dump di e-mail	145
Il protocollo di posta elettronica e la struttura dei messaggi.....	146
Formati di file per dump di posta elettronica	147
File EML.....	147
File MBOX.....	148
File di dati PST di Outlook	148
Esercizio 6.1 – Scaricare dump di posta elettronica da tre dataset.....	149
Il dataset delle forze di polizia di Nauru	149
Il dataset Oath Keepers	149
Il dataset di Heritage Foundation	150
Ricerca nei dump di e-mail con Thunderbird	150
Esercizio 6.2 – Configurare Thunderbird per i dump di e-mail....	151
Lettura di singoli file EML con Thunderbird	152
Esercizio 6.3 – Importare il dump di e-mail EML delle forze di polizia di Nauru.....	153
Ricerche nella posta elettronica in Thunderbird.....	154
Ricerche con Filtro veloce.....	155
La finestra di dialogo Ricerca messaggi	155
Esercizio 6.4 – Importare il dump di e-mail MBOX di Oath Keepers.....	155
Esercizio 6.5 – Importare il dump di e-mail PST di Heritage Foundation	157
Altri strumenti per la ricerca nei dump di e-mail.....	159
Microsoft Outlook.....	159
Aleph	161
Riepilogo	162
Parte III Programmazione Python	163
Capitolo 7 Introduzione a Python.....	165
Esercizio 7.1 – Installare Python.....	165
Windows.....	166
Linux	166
Mac OS.....	166
Esercizio 7.2 – Sviluppare uno script Python.....	166

Nozioni di base di Python	168
L'interprete interattivo di Python	168
Commenti	168
Calcoli con Python	169
Stringhe	171
Esercizio 7.3 – Sviluppare uno script Python con variabili, calcoli e stringhe	172
Liste e cicli	173
Definizione e visualizzazione delle liste	174
Cicli	176
Flusso di controllo	177
Operatori di confronto	178
Istruzioni if	179
Blocchi di codice annidati	180
Ricerca nelle liste	181
Operatori logici	181
Gestione delle eccezioni	183
Esercizio 7.4 – Cicli e flusso di controllo	185
Funzioni	187
La parola chiave def	188
Argomenti di default	189
Valore restituito	190
Docstring	191
Esercizio 7.5 – Esercitazioni sulla scrittura delle funzioni	192
Riepilogo	193
Capitolo 8 Elaborazione dei dati in Python	195
Moduli	195
Template per script Python	197
Esercizio 8.1 – Esplorare i file contenuti in BlueLeaks	198
Elencare i file contenuti in una cartella	198
Contare i file e le cartelle contenute in una cartella	199
Esplorazione delle cartelle con <code>os.walk()</code>	201
Esercizio 8.2 – Trovare i file più grandi di BlueLeaks	202
Moduli di terze parti	203
Esercizio 8.3 – Esercitarsi sugli argomenti della riga di comando con Click	206
Gli argomenti della riga di comando evitano di ricorrere all'hardcoding	208
Esercizio 8.4 – Trovare i file più grandi in qualsiasi dataset	208
Dizionari	209
Definizione dei dizionari	210
Estrazione e impostazione dei valori	210
Navigazione nei dizionari e nelle liste dei registri delle chat di Conti	211

Esplorazione di dizionari e liste di dati in Python.....	212
Selezione di valori nei dizionari e nelle liste.....	214
Analisi dei dati archiviati in dizionari e liste.....	215
Esercizio 8.5 – Mappare i file CSV in BlueLeaks.....	218
Accettare un argomento dalla riga di comando.....	219
Creare cicli sulle cartelle di BlueLeaks.....	220
Compilare il dizionario.....	221
Visualizzare l’output.....	223
Letture e scrittura di file.....	224
Apertura dei file.....	224
Scrittura di righe su un file.....	225
Letture di righe da un file.....	226
Esercizio 8.6 – Esercitarsi a leggere e scrivere file.....	227
Riepilogo.....	229

Parte IV Dati strutturati231

Capitolo 9 Blueleaks, Black Lives Matter e il formato CSV.....233

Installazione del software per fogli di lavoro.....	233
Il formato CSV.....	234
Esplorazione di file CSV con software per fogli di lavoro e editor di testo.....	236
La mia indagine su BlueLeaks.....	239
Concentrarsi su un fusion center.....	239
Che cos’è l’NCRIC.....	239
Indagine su un SAR.....	240
Letture e scrittura di file CSV in Python.....	244
Esercizio 9.1 – Rendere più leggibili i file CSV del dataset BlueLeaks.....	245
Accettare il percorso CSV come argomento.....	246
Creare cicli sulle righe del file CSV.....	246
Visualizzare i campi del file CSV su righe distinte.....	248
Letture delle e-mail dai fusion center.....	249
Elenchi di dimostrazioni dei Black Lives Matter.....	250
Memo di “intelligence” dall’FBI e dal DHS.....	254
Breve manuale di HTML.....	255
Esercizio 9.2 – Rendere più leggibili le e-mail.....	257
Accettare gli argomenti della riga di comando.....	257
Creare la cartella di output.....	258
Definire il nome di file per ogni riga.....	259
Scrivere la versione HTML di ogni e-mail.....	260
Individuazione del nome e dell’URL dei siti considerati da BlueLeaks.....	264
Esercizio 9.3 – Creare un file CSV dei siti considerati da BlueLeaks.....	266

Aprire in scrittura un file CSV	267
Trovare tutti i file Company.csv	268
Aggiungere al file CSV i siti considerati da BlueLeaks	269
Riepilogo	271

Capitolo 10 BlueLeaks Explorer273

Rivelazioni da scoprire in BlueLeaks	274
Esercizio 10.1 – Installare BlueLeaks Explorer	274
Creare il file di configurazione di Docker Compose	275
Attivare il container	276
Inizializzare i database	276
La struttura dell’NCRIC.....	278
Esplorazione delle tabelle e delle relazioni	278
Ricerca per parole chiave	281
Costruzione della struttura di BlueLeaks	281
Definizione della struttura del JRIC.....	282
Visualizzazione dei campi utili	284
Modifica del tipo dei campi	286
Aggiunta della tabella Leads del JRIC	287
Costruzione di una relazione	288
Verifica dei dati di BlueLeaks	291
Esercizio 10.2 – Completare la costruzione della struttura del JRIC.....	292
La tecnologia su cui si basa BlueLeaks Explorer	293
Il back-end	293
Il frontend	293
Riepilogo	294

Capitolo 11 Parler, l’insurrezione del 6 gennaio e il formato JSON295

Le origini del dataset Parler.....	296
Come sono stati archiviati i video di Parler	296
L’impatto del dataset sul secondo impeachment di Trump.....	297
Esercizio 11.1 – Scaricare ed estrarre i metadati dai video di Parler.....	298
Scaricare i metadati.....	298
Decomprimere e scaricare i singoli video di Parler	299
Estrarre i metadati di Parler.....	301
Il formato JSON.....	302
La sintassi di JSON	303
Parsing dei dati JSON con Python.....	305
Gestione delle eccezioni con JSON	308
Strumenti per esplorare i dati JSON.....	309
Conteggio dei video dotati di coordinate GPS utilizzando grep	309

Formattazione e ricerca dei dati con il comando jq	310
Esercizio 11.2 – Sviluppare uno script per filtrare i video con GPS del 6 gennaio 2021	311
Accettare come argomento il percorso dei metadati di Parler	312
Scorrere i file di metadati di Parler	313
Filtrare i video dotati di coordinate GPS	314
Filtrare i video del 6 gennaio 2021	316
Gestione delle coordinate GPS	317
Ricerca per latitudine e longitudine	317
Conversione tra i vari formati di coordinate GPS	319
Calcolo della distanza GPS in Python	320
Ricerca del centro di Washington	322
Esercizio 11.3 – Aggiornare lo script per filtrare i video sull'insurrezione	322
Tracciamento delle coordinate GPS su una mappa con simplekml	325
Esercizio 11.4 – Creare file KML per visualizzare i dati di localizzazione	327
Creare un file KML per tutti i video dotati di coordinate GPS	328
Crea file KML per i video dal 6 gennaio 2021	331
Visualizzazione dei dati di localizzazione con Google Earth	333
Visualizzazione dei metadati con ExifTool	336
Riepilogo	338

Capitolo 12 Epik Fail, indagini sull'estremismo e database SQL.....339

La struttura dei database SQL	340
Database relazionali	341
Client e server	342
Tabelle, colonne e tipi	342
Esercizio 12.1 – Creare e sottoporre a test un server MySQL utilizzando Docker e Adminer	344
Utilizzare il server	344
Connettersi al database con Adminer	345
Creare un database di test	346
Esercizio 12.2 – Interrogare il database SQL	347
Istruzioni INSERT	347
Istruzioni SELECT	349
Clausole JOIN	354
Istruzioni UPDATE	357
Istruzioni DELETE	357
Introduzione al client a riga di comando MySQL	357
Esercizio 12.3 – Installare e sottoporre a test il client MySQL a riga di comando	359

Query specifiche per MySQL	360
La storia di Epik	362
L'hacking di Epik	363
I dati WHOIS di Epik	364
Esercizio 12.4 – Scaricare ed estrarre parte del dataset Epik	366
Esercizio 12.5 – Importare dati del dataset Epik in MySQL	367
Creare un database per api_system	367
Importare i dati di api_system	367
Esplorazione del database SQL di Epik	369
La tabella domain	369
La tabella privacy	371
Le tabelle hosting e hosting_server	373
Utilizzo dei dati di Epik nel cloud	375
Riepilogo	377

Parte V Casi di studio379

Capitolo 13 Profittatori pandemici e disinformazione sul Covid-19381

Le origini di AFLDS	382
I dataset Cadence Health e Ravkoo	385
Estrazione dei dati in un container di file crittografati	385
Analisi dei dati con strumenti a riga di comando	386
Creazione di un foglio di lavoro dei pazienti	393
Calcolo delle entrate per le prescrizioni mediche di Ravkoo	397
Ricerca del prezzo e della quantità dei farmaci venduti	397
Categorizzazione per farmaco dei dati delle prescrizioni	400
Uno sguardo più approfondito sui dati dei pazienti di Cadence Health	402
Ricerca dei partner di Cadence Health	402
Ricerca dei pazienti per città	405
Ricerca dei pazienti per età	409
Autenticazione dei dati	413
Le conseguenze	414
Regole di notifica delle violazioni dell'HIPAA	415
Inchiesta del Congresso	415
La nuova avventura imprenditoriale di Simone Gold	416
Scandalo e lotte interne in AFLDS	416
Riepilogo	417

Capitolo 14 I neonazisti e le loro chatroom419

Come gli antifascisti si sono infiltrati nei server neonazisti di Discord	420
--	-----

Analisi dei registri delle chat trapelati	421
Come rendere leggibili i file JSON	422
Esplorazione di oggetti, chiavi e valori con jq	423
Conversione di timestamp	428
Ricerca degli utenti	428
Il tracker della cronologia di Discord	430
Uno script per cercare nei file JSON	432
Il mio codice Discord Analysis	436
Progettazione del database SQL	437
Importazione dei registri delle chat nel database SQL	440
Costruzione dell'interfaccia web	445
Uso di Discord Analysis per trovare rivelazioni	451
Il server Discord Pony Power	453
Il lancio di DiscordLeaks	457
Gli sviluppi	457
La causa contro Unite the Right	458
I registri della chat di Patriot Front	458
Riepilogo	460
Epilogo	461
Appendice A Soluzioni ai problemi più comuni di WSL	463
Il filesystem Linux di WSL	464
Il problema delle prestazioni del disco	466
Soluzioni al problema delle prestazioni del disco	467
Memorizzazione in Linux dei soli dataset attivi	467
Memorizzazione del filesystem Linux su un disco USB	467
Prossimi passi	472
Appendice B Scraping del Web	473
Considerazioni legali	474
Le richieste HTTP	474
Tecniche di scraping	475
Caricamento di pagine con HTTPX	476
Parsing del codice HTML con BeautifulSoup	481
Automazione dei browser web con Selenium	486
Prossimi passi	491
Indice analitico	493