HSLU Hochschule Luzern

CF: Tools

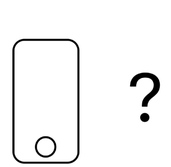
CF HS24

Dr. Hannes Spichiger

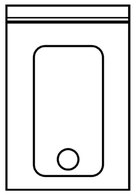
Departement Informatik
21.09.2023

FH Zentralschweiz

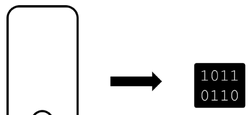
Forensischer Prozess: Welche Schritte können wir automatisieren?



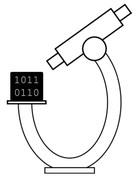
Recognition



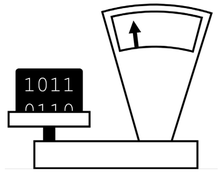
Preservation



Acquisition



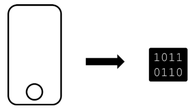
Examination

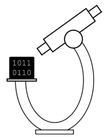


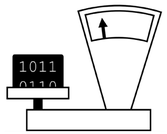
Interpretation

Reporting

Datenverarbeitungs-Pipeline







Disk → Acquisition → Rohdaten → Aufbereitung → Dargestellte Daten → Analyse → Resultate

- Rohdaten: Disk Image
- Dargestellte Daten: Datei-System, Dateien
- Resultate: Multimedia, Zeitstrahl, Karte, Chatverläufe, Geräteinfo

HSLU Hochschule Luzern

Acquisition-Funktionalitäten

11.10.2024

FH Zentralschweiz

Acquistion-Funktionalitäten

- Lesen von Datenträgern in einem Read-Only-Modus
- Erstellen von Disk-Images
- Live-Collection
- Erstellen und überprüfen von Hashes

Acquistion-Fokussierte Tools

exterro

FTK® Imager

Cellebrite DIGITAL COLLECTOR

HSLU Hochschule Luzern

Aufbereitungs-Funktionalitäten

11.10.2024

FH Zentralschweiz

HEX-Ansicht

- Im Grundsatz ist ein Datensatz vor allem erstmal eine Bit-Folge
- Wir können alle Daten als diese Bit-Folge darstellen
- Typischerweise wird eine Hex-Ansicht gewählt
  - 2 Stellen Hex = 1 Byte
- Eg. HxD

File-System Ansicht

- Disk-Image enthält meist Partitionen mit Dateisystemen

- Diese können Einfach als Ordner-Struktur dargestellt werden
  - Partitionen
    - Ordner
    - Files

- Manche Tools Fokussieren Stark auf diese Sichtweise
  - XWays
  - FTK
  - Encase
  - (Autopsy)

X-Ways

exterro

FTK® Forensic Toolkit

EnCase Forensic Suite

AUTOPSY
DIGITAL FORENSICS

File-System Ansicht: Wie Stelle ich Elemente dar, die kein File sind?

- Nicht alles ist ein File:
  - Inter-Partition Space
  - Raum zwischen Dateien
  - File-Slack

- Wird typischerweise in die «File-Logik» gepresst
  - Inter-Partition-Space => Wird als eigene Partition dargestellt
  - Raum zwischen Dateien => Wird als Datei Dargestellt

- Wo in der Ordner-Struktur?
  - Virtueller Ordner mit «Dateien» die nicht in einem effektiven Ordner liegen

HSLU Hochschule Luzern

Analyse-Funktionalitäten

11.10.2024

FH Zentralschweiz

Event-Fokussierte Ansicht

- Jedes Element im Datenset wird als Event betrachtet

- Bedingt Parsen der Daten vor Ansicht
  - Vor Parsing könnt ihr Daten meist nicht einmal betrachten

- Event-Fokussierte Tools
  - Magnet
  - Nuix
  - Cellebrite Inspector
  - (Autopsy)

- Event-Basierte Analyse ist oft stärker OS-Abhängig

Parsen von Datenstrukturen

- Informationen werden oft in anderen Datenstrukturen abgelegt
  - Datenbanken (Eg. Browser-History)
  - Log-Strukturen (Eg. User-Logins)
  - Registry-Hives (Eg. Systemeinstellungen)

- Oft ist die Datenstruktur weniger interessant als die Einträge

- Tools bieten oft Funktionalitäten an, um diese Strukturen zu Parsen

- Das Tool erstellt pro Eintrag ein Objekt, welches die geparsten Informationen enthält

Voraussetzungen für Daten-Parsing

- Datenstruktur ist dem Tool bekannt
- Datenstruktur stimmt mit der bekannten Struktur überein
- Datenstruktur befindet sich am Ort, der vom Tool erwartet wird.
- Eg. Ein Tool wird die Browserhistory nur parsen können, wenn
  - Der Browser bekannt ist
  - Die Version eine bekannte Struktur enthält
  - Der Browser nicht in einer seltsamen Ordner-Struktur installiert wurde

Parsen von System-Informationen

- Die Menge an Verfügbaren Systeminformationen ist enorm
- Tool-Hersteller müssen eine Auswahl treffen
- Nicht alle Tools treffen dieselbe Auswahl
- Nur weil ein Tool eine Information nicht geparst präsentiert, heisst das nicht, dass sie nicht da ist!

Unpacking

- Dateisysteme können komprimierte Daten enthalten

- Damit ihr diese Lesen könnt, müsst ihr sie dekomprimieren können

- Viele Tools bieten Unpacking-Funktionalitäten an
  - Zip-Files
  - Embedded Files (Eg. Bilder in PDF)
  - Embedded Thumbnails

- Das Tool stellt (hoffentlich) sicher, dass nachvollziehbar ist, aus welcher Datei die extrahierten Dateien stammen
  - Oft als «Child» der Ursprungsdatei dargestellt
  - In Ordner-Ansicht wird «Parent»-Datei zu einem Ordner, der die «Child»-Dateien enthält

Timelineing

- Chronologische Darstellung aller Ereignisse

- Die Meisten Tools bieten irgendeine solche Funktionalität an

- Oft nicht besonders Optimiert
  - Viele Daten auf einmal
  - Meist langsam

- Grosse Schwierigkeit: Welchen Timestamp verwende ich?
  - File hat 3 / 4 Timestamps
  - Tools, welche nicht auf Timelineing optimiert sind, lösen das problem oft nicht besonders gut

Timelineing-Tools

https://github.com/log2timeline

splunk>

Stack

Multimedia-Analyse

- Bild- & Video-Triage gehört im LE-Kontext zu häufigen Herausforderungen

- Viele Tools bieten Möglichkeiten, um Bildanalysen durchzuführen
  - Skin-Tone
  - Nudity-Detection
  - Content-Detection
  - Still-Image-Capture (für Video)

- Details in der Multimedia-Vorlesung

Statistische Analysen

- Daten haben eine bestimmte Struktur
- Nicht alle Byte-Werte kommen gleich häufig vor
- Anders bei Komprimierten oder Verschlüsselten Daten
  - Manche Tools bieten Funktionen zur Statistischen Analyse an, um solche Dateien zu finden
- Statistische Analysen können auch manche Steganographie-Ansätze identifizieren

Multi-Source-Analyse

- Fälle werden immer komplexer

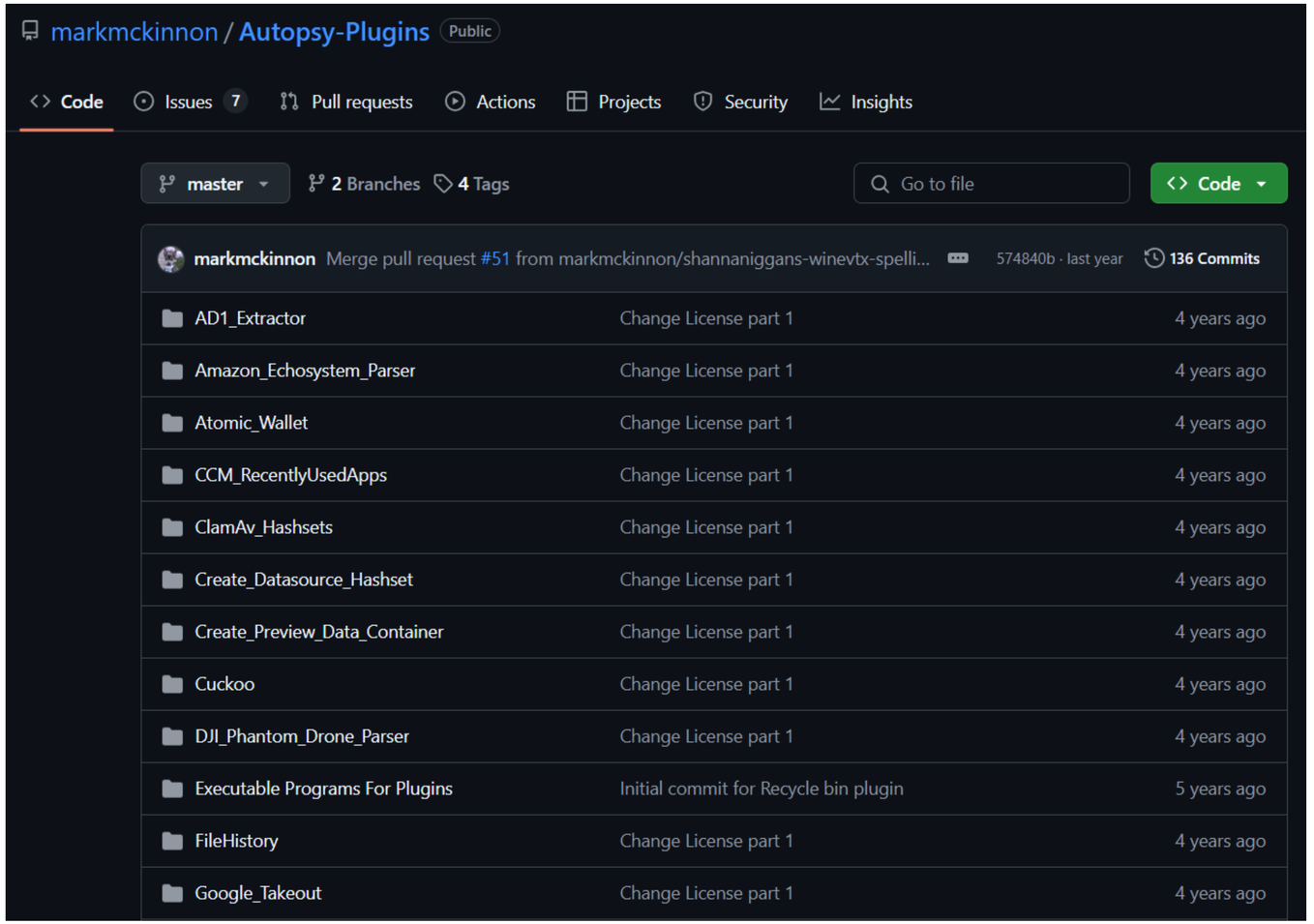
- Meist habt ihr mehrere Datenquellen

- Funktionalitäten, um Daten aus mehreren Quellen untereinander abzugleichen sind derzeit sehr in
  - Dateien, die mehrmals auftauchen
  - Gemeinsame Kontakte

- Ich bin eher skeptisch, wie gut diese Funktionalitäten implementiert sind.
  - Ich habe noch nie eine Demo mit mehr als 3 Geräten in derselben Analyse gesehen

Plugins

- Manche Software erlaubt das ausführen von externen Plugins
  - Primär Autopsy
- Ermöglicht Analyse von spezifischeren Anwendungen / Datenquellen
- Forschungsresultate werden oft als Autopsy-Plugin zur Verfügung gestellt



HSLU Hochschule Luzern

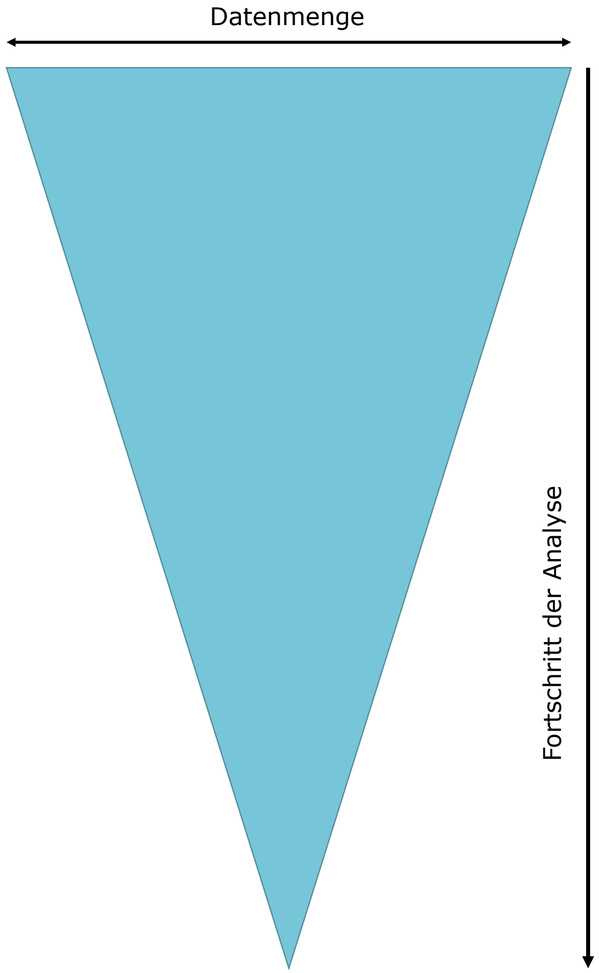
Triage

11.10.2024

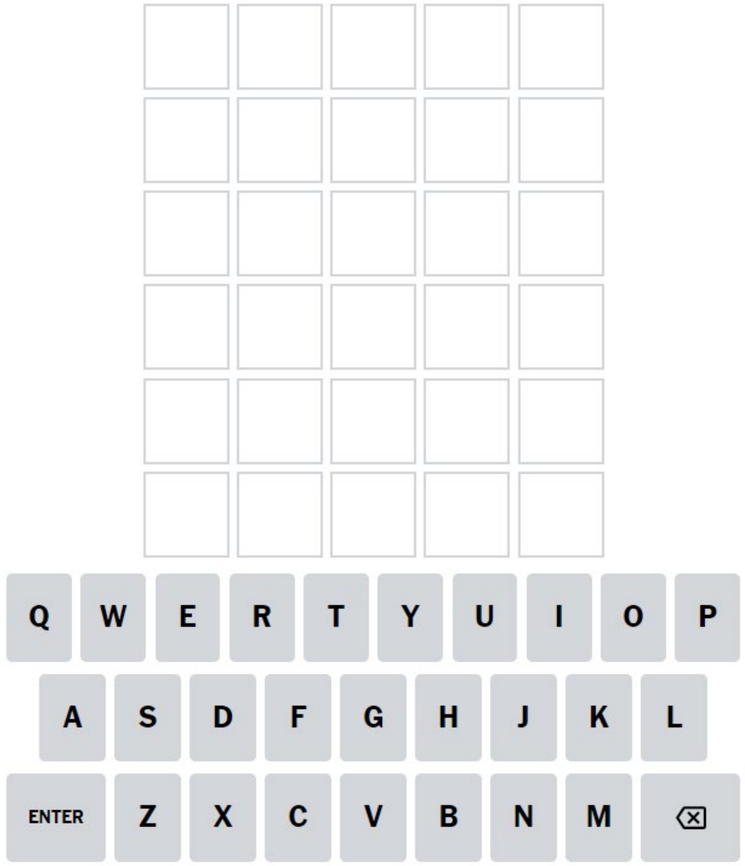
FH Zentralschweiz

Triage: Reduzieren der Datenmenge

- Zu Beginn der Analyse haben wir alle Möglichen Daten
- Die meisten davon sind nicht Fallrelevant
- Wie können wir diese Menge reduzieren?



WORDLE: Wie gehen wir optimal vor?



Wie können wir die restlichen Antworten möglichst effizient reduzieren?

- Wir haben einen initialen Pool von Antworten
  - Wir suchen nach Tests, welche möglichst effizient unseren Pool reduzieren

- Mass für «Effizienz» in diesem Kontext: Power of Discrimination (PD)
  - $PD = 1 - \sum_i {G_i}^2$
  - G_i: Häufigkeit des beobachteten Resultats i beim Anwenden des Tests auf die Zielgruppe
  - $\sum_i G_i = 1$

- Z.B: Erster Buchstabe des Wortes = A
  - $PD = 1 - {G_{1=A}}^2 - {G_{1\neq A}}^2$

- Wir können zwischen diversen Testmethoden diejenige mit der höchsten PD suchen.
  - Optimaler Fall: Restgruppen sind für alle Resultate gleich gross
  - $PD_{Max(2)} = 1 - 0.5^2 - 0.5^2 = 0.5$

Wordle: Erster Buchstabe

i	#(1 = i)	#(1≠i)	PD
A	140	2169	0.114
B	173	2136	0.139
C	198	2111	0.157
D	111	2198	0.092
E	72	2237	0.060
F	135	2174	0.110
G	115	2194	0.095
H	69	2240	0.058
I	34	2275	0.029
J	20	2289	0.017
K	20	2289	0.017
L	87	2222	0.073
M	107	2202	0.088

i	#(1 = i)	#(1≠i)	PD
N	37	2272	0.032
O	41	2268	0.035
P	141	2168	0.115
Q	23	2286	0.020
R	105	2204	0.087
S	365	1944	0.266
T	149	2160	0.121
U	33	2276	0.028
V	43	2266	0.037
W	82	2227	0.069
X	0	2309	0.000
Y	6	2303	0.005
Z	3	2306	0.003

WORDLE

- Wordle-Eingaben sind komplexer als «Erster Buchstabe = A»
  - Wir müssen ein Wort aus der Liste Testen
  - Pro Position gibt es 3 Mögliche Antworten
  - Pro Eingabe können wir 125 verschiedene Antworten erhalten

- Die Beste Strategie ist, jeweils das Wort zu wählen, welches die Restlichen möglichen Lösungen am besten über die möglichen Antworten verteilt.

Mögliche Resultate für Start mit “TRACE”

pattern	count	guess
*****	246	biddy
****Y	123	beefy
****G	104	belie
***Y*	48	child
***YY	13	bicep
***YG	16	chide
***G*	38	block
***GY	10	beech
***GG	15	deuce
Y	127	abyss
**Y*Y	48	abbey
**Y*G	45	abide
**YY*	32	bacon
**YYY	10	cagey
**YYG	4	cable
**YG*	2	fancy
**YGY	1	mecca
**YGG	3	dance
G	49	again
**G*Y	8	beady
**G*G	35	abase
**GY*	17	chaff

pattern	count	guess
**GYY	0	
**GYG	4	cease
**GG*	12	aback
**GGY	3	beach
**GGG	3	peace
*Y***	64	blurb
*Y**Y	113	berry
*Y**G	40	borne
*Y*Y*	19	chirp
*Y*YY	14	cheer
*Y*YG	5	chore
*Y*G*	3	birch
*Y*GY	2	mercy
*Y*GG	1	force
*YY**	61	abhor
*YY*Y	41	aider
*YY*G	9	adore
YYY	13	acorn
*YYYY	5	caper
*YYYG	1	carve
YYG	3	circa
*YYGY	0	

pattern	count	guess
*YYGG	1	farce
*YG**	17	alarm
*YG*Y	9	beard
*YG*G	7	aware
YGY	5	chair
*YGYY	0	
*YGYG	1	scare
YGG	1	roach
*YGGY	1	reach
*YGGG	0	
*G***	49	bring
*G**Y	19	breed
*G**G	23	bribe
*G*Y*	11	crimp
*G*YY	8	credo
*G*YG	5	creme
*G*G*	5	brick
*G*GY	1	wreck
*G*GG	1	price
*GY**	13	arbor
*GY*Y	7	arena
*GY*G	3	argue

pattern	count	guess
GYY	1	croak
*GYYY	2	creak
*GYYG	0	
GYG	0	
*GYGY	0	
*GYGG	0	
*GG**	25	braid
*GG*Y	0	
*GG*G	10	brake
GGY	6	cramp
*GGYY	0	
*GGYG	3	crane
GGG	2	crack
*GGGY	0	
*GGGG	2	brace
Y****	113	bigot
Y***Y	58	befit
Y***G	20	butte
Y**Y*	8	cloth
Y**YY	7	chest
Y**YG	2	chute
Y**G*	12	botch

pattern	count	guess
Y**GY	5	edict
Y**GG	0	
Y*Y**	53	abbot
Y*Y*Y	19	adept
Y*Y*G	10	atone
Y*YY*	8	antic
Y*YYY	5	cadet
Y*YYG	2	acute
Y*YG*	7	batch
Y*YGY	0	
Y*YGG	0	
Y*G**	21	adapt
Y*G*Y	12	beast
Y*G*G	12	abate
Y*GY*	3	chant
Y*GYY	0	
Y*GYG	0	
Y*GG*	1	stack
Y*GGY	2	enact
Y*GGG	0	
YY***	32	birth
YY**Y	29	beret

pattern	count	guess
YY**G	3	forte
YY*Y*	1	court
YY*YY	1	recut
YY*YG	0	
YY*G*	0	
YYY**	12	abort
YYY*Y	10	after
YYY*G	0	
YYYY*	2	actor
YYYYY	1	cater
YYYYG	0	
YYYG*	0	
YYYGY	0	
YYYGG	0	
YYG**	7	apart
YYG*Y	1	heart
YYG*G	1	stare
YYGY*	1	chart
YYGYY	0	
YYGYG	0	

pattern	count	guess
YYGG*	0	
YYGGY	1	react
YYGGG	0	
YG***	14	broth
YG**Y	3	erupt
YG**G	3	brute
YG*Y*	2	crust
YG*YY	2	crept
YG*YG	0	
YG*G*	0	
YG*GY	1	erect
YG*GG	0	
YGY**	1	artsy
YGY*Y	1	great
YGY*G	0	
YGYY*	0	
YGYYY	0	
YGYYG	0	
YGYG*	0	
YGYGY	0	
YGYGG	0	
YGG**	4	draft

pattern	count	guess
YGG*Y	0	
YGG*G	2	grate
YGGY*	1	craft
YGGYY	0	
YGGYG	1	crate
YGGG*	0	
YGGGY	0	
YGGGG	0	
G****	17	thigh
G***Y	14	teddy
G***G	11	tense
G**Y*	4	tonic
G**YY	0	
G**YG	0	
G**G*	2	thick
G**GY	0	
G**GG	1	twice
G*Y**	21	tabby
G*Y*Y	3	taken
G*Y*G	2	table
G*YY*	2	tacit
G*YYY	0	

pattern	count	guess
G*YYG	0	
G*YG*	0	
G*YGY	0	
G*YGG	0	
G*G**	3	thank
G*G*Y	0	
G*G*G	1	tease
G*GY*	0	
G*GYY	0	
G*GYG	0	
G*GG*	0	
G*GGY	1	teach
G*GGG	0	
GY***	11	third
GY**Y	7	tenor
GY**G	3	terse
GY*Y*	0	
GY*YY	0	
GY*YG	0	
GY*G*	1	torch
GY*GY	0	
GY*GG	0	

pattern	count	guess
GYY**	3	tapir
GYY*Y	4	taker
GYY*G	0	
GYYY*	0	
GYYYY	0	
GYYYG	0	
GYYG*	0	
GYYGY	0	
GYYGG	0	
GYG**	1	tiara
GYG*Y	1	teary
GYG*G	0	
GYGY*	0	
GYGYY	0	
GYGYG	0	
GYGG*	0	
GYGGY	0	
GYGGG	0	
GG***	10	troll
GG**Y	3	trend
GG**G	5	tribe
GG*Y*	0	

pattern	count	guess
GG*YY	0	
GG*YG	0	
GG*G*	2	trick
GG*GY	0	
GG*GG	2	trice
GGY**	2	triad
GGY*Y	2	tread
GGY*G	0	
GGYY*	0	
GGYYY	0	
GGYYG	0	
GGYG*	0	
GGYGY	0	
GGYGG	0	
GGG**	6	trail
GGG*Y	0	
GGG*G	1	trade
GGGY*	0	
GGGYY	0	
GGGYG	0	
GGGG*	2	track
GGGGY	0	
GGGGG	1	trace

Quelle: https://real-statistics.com/wordle-strategy/winning-wordle-in-two-tries/

Situation in der Computer-Forensik

- Wir suchen nach relevanten Informationen zu einer Situation
- Wir haben eine Grosse Menge an Daten
- Wir wollen Filter auf die Daten anwenden, welche die Datenmenge, welche wir von Hand analysieren müssen reduziert
- Anders als bei WORDLE haben wir nicht ein Frage-Antwort Spiel
  - Aus der WORDLE-Logik können wir die Schrittweise Herangehensweise mitnehmen

Filter-Kriterien

- Ein gutes Filter-Kriterium
  - Hat eine hohe Wahrscheinlichkeit, irrelevante Daten auszuschliessen
  - Hat eine sehr tiefe Wahrscheinlichkeit, relevante Daten auszuschliessen

- Was weiss ich über meine Daten?
  - Zeitraum
  - Dateityp
  - Dateigrösse
  - Analyse-Resultat
  - Speicherort

- Nach welchen Kriterien kann ich filtern?

Nach dem Filtern: Sortieren

- Manchmal ist es nicht möglich einen Cut-Off zu definieren

- In diesen fällen können Dateien nach einem Kriterium geordnet werden
  - Manuelle Analyse in Reihenfolge der Liste folgt der Logik, je weiter unten, desto weniger wahrscheinlich relevant

- Z.B. Fragestellung: Hat eine Person selber illegales Bildmaterial produziert?
  - Dateien, die übers Netz geteilt werden sind typischerweise komprimiert
  - Grössere Dateien sind wahrscheinlicher von dem Nutzer selber produziert

⇒ Filtern nach Bild-Dateien

⇒ Sortieren nach Dateigrösse von gross nach klein

⇒ Selbstgemachte Bilder schwimmen oben auf

Key-Word Suche

- Suchen nach Stichworten

- Meist zyklisch
  - Was erwarte ich?
  - Anpassen der Suchbegriffe aufgrund der Resultate

- Bedingt Indexierung der Daten
  - Teilweise einzelner Schritt

- Welche Daten werden alles indexiert?
  - Resultate oder Rohdaten?
  - In welchen Bereichen?
  - Komprimierte Daten?
  - Datenformate (Nur ASCII oder auch UTF?)

⇒ XWays ist für Textsuchen wahnsinnig mächtig, weil es nahezu alles indexiert
⇒ HxD ist sehr mächtig, weil es sehr viele verschiedene Encoding-Formen unterstützt

Hash-Suche

- Suchen nach bekannten Dateien

- Known Irrelevant
  - Eg. Systemdateien
  - NIST Software Reference Library (NISRL)
  - https://www.nist.gov/itl/ssd/software-quality-group/national-software-reference-library-nsrl

- Known Relevant
  - Malware-Datenbanken
  - Files, die ihr erwartet

Beispiel: DFRWS EU 2021 Rodeo

- Ihr erhaltet ein Disk-Image
  - Entweder, das Gerät, von dem das Image stammt war in einem definierten Zeitraum auf einer bestimmten Website und jemand hat versucht, nachträglich die Spuren davon zu verwischen
  - Oder, das Gerät war nie auf dieser Website

- Rodeo-Setting: in 90’ sollt ihr 10 solche Images kategorisieren

- Überlegt euch eine Vorgehensweise, wie ihr das schaffen könnt

The Fictitious Task

You are given the disk image of Markus Maier’s laptop. The police requests answers to the following two questions:

1. Has the website https://de.wikipedia.org/wiki/Nashörner been accessed between October 20 and November 5, 2019?
2. Were pictures downloaded from https://de.wikipedia.org/wiki/Nashörner onto the computer between October 20 and November 5, 2019?

Because of the situation, there might be the possibility of evidence tampering (i.e., a post-mortem manipulation of the main disk image).

https://www.cybercrime.fau.de/dfrws-eu-2021-forensic-rodeo/

HSLU Hochschule Luzern

Allgemeine Funktionalitäten

11.10.2024

FH Zentralschweiz

Dokumentation

- Was wurde gemacht?
  - E.G. Acquisition-Log by Kopier-Tools

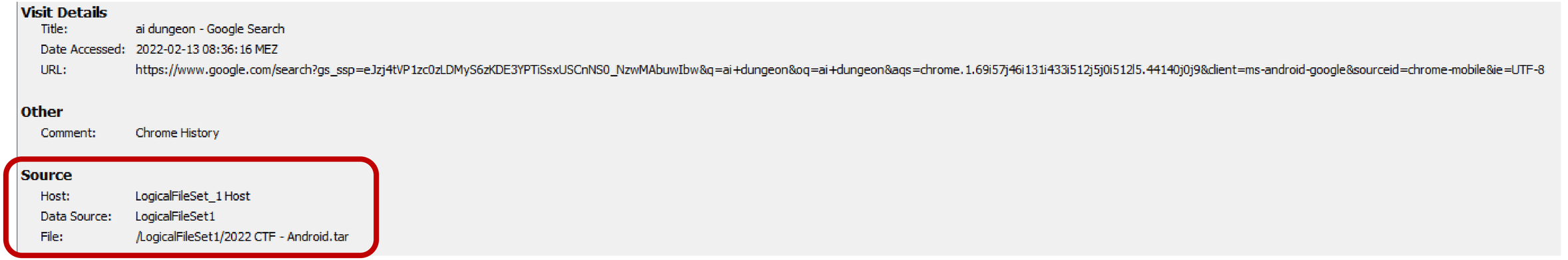
- Auch in Tools später in der Pipeline interessant
  - Welche Plugins wurden eingesetzt?
  - Welche Einstellungen wurden bei einer Analyse verwendet?
  - In welcher Reihenfolge?

- Tools stellen mehr oder weniger fortgeschrittene Lösungen dafür zur Verfügung
  - Logs
  - Screenshots

Nachvollziehbarkeit

- Woher kommt eine Information?
  - Welche Datei / Datenbank-Eintrag / Registry
  - Von welcher Funktion gefunden

- Erlaubt mir zu überprüfen, ob das Resultat stimmt



Kompatibilität

- Wenn ich mehrere Tools verwende, insb. in einer Pipeline, müssen Export-Formate von späteren Tools gelesen werden können

- Bei Disk-Images haben wir Standard-Formate
  - Raw
  - E01

- Sonst noch sehr schwierig

Erstellen von „Reports“

- Viele Tools ermöglichen das erstellen von „Reports“
- Dies sind nicht Berichte im eigentlichen Sinn
- Es handelt sich eher um Resultat-Exporte
- Qualität und Nutzen dieser Reports ist meistens zwischen mässig und katastrophal angesiedelt

Kompatibilität: Export von Dateien

- Nahezu alle Tools erlauben einen Export von Dateien
- Diese können dann in externen Tools angeschaut oder analysiert werden
- Durch den Export gehen sämtliche Kontext-Informationen verloren

Standardisierungs-Versuche: CASE

- Ziel: Einheitliches Framework um Daten und Resultate Darzustellen
- Würde u.A. Interoperabilität zwischen Tools ermöglichen
- Projekt ist noch nicht so weit fortgeschritten, dass Implementierung schon Standard wäre

https://caseontology.org

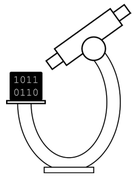
HSLU Hochschule Luzern

Zuverlässigkeit von Tools

11.10.2024

FH Zentralschweiz

Zuverlässigkeit von Tools



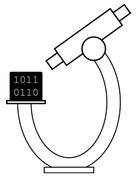
- Inwiefern sind wir zuversichtlich, dass unser Tool die erwarteten Resultate liefert?

- Hängt vom Tool ab
- Hängt von der Komplexität der Aufgabe ab

⇒ Allgemein wird empfohlen, eine Analyse mit verschiedenen Tools durchzuführen

⇒ In der Praxis macht das kaum jemand

Zuverlässigkeit von Tools



- Resultate, welche vorkommen sollten, aber nicht da sind
  - Kann bedeuten, dass wir Spuren verpassen

- Resultate, welche auftauchen, aber nicht dem entsprechen, was wir suchen
  - Bedeutet vor allem mehr Arbeit

- Resultate welche mehrmals genannt werden
  - Sind dies eigenständige Resultate oder dieselbe Information aus verschiedenen Quellen?

- Resultate, die falsch sind
  - Manchmal nachvollziehbar
  - Manchmal nicht

HSLU Hochschule Luzern

Anschaffen von Tools

11.10.2024

FH Zentralschweiz

Was brauche ich?

- Definiert aufgrund eurer Aufträge Anforderungen
- Was für Fälle habe ich häufig?
- Was für Spurentypen / Datentypen sind dort von Interesse?
- Welche Fragestellungen muss ich häufig beantworten?

Testen

- Lasst euch Test-Lizenzen geben
- Probiert die Funktionalitäten aus
- Challenge-Datensets sind oft interessant, um Tools auszuprobieren
- Macht das Tool das, was ihr wollt?
- Ist das Tool in der Handhabung praktisch?
- Ist das Tool kompatibel mit anderen Lösungen, die ihr bereits verwendet?

Open-Source vs. Kommerziell

Open Source

- Prozesse sind Nachvollziehbar
- Community-Driven
  - Decken z.T. auch sehr spezifische Fälle ab
- Updates nicht garantiert
- Meist nicht optimiert

Kommerziell

- Prozesse nicht immer transparent
- Neuentwicklungen abhängig von dem, was oft verlangt wird
- Meist aktueller als Open Source-Lösungen
- In spezifischen Problemen oft mächtiger & effizienter als Open Source-Lösungen

Literatur empfiehlt Open Source-Lösungen, In der Praxis sind kommerzielle Tools oft unumgänglich

HSLU Hochschule Luzern

Fragen?

Hochschule Luzern
Informatik
Dr. Hannes Spichiger
Dozent

T direkt +41 41 349 31 24
Hannes.spichiger@hslu.ch

Auftrag auf nächste Woche

- Stellt sicher, dass ihr ein funktionierendes Analyse-Tool habt

- E.G. Autopsy (https://www.sleuthkit.org/autopsy/)