ORKL

Eine Bibliothek für Cyber Threat Intelligence

Robert Haist - 30. DFN-Konferenz 2023

$WHOAMI

Robert Haist

CISO | VP Security @ TeamViewer

M.Sc. Advanced Security and Digital Forensics Edinburgh Napier University

Master Thesis:

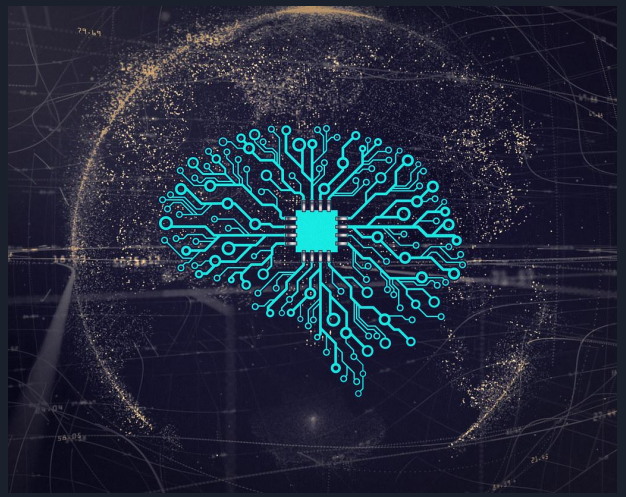
"*TIRAKL: an NLP assisted approach to curate OSINT Cyber Threat Intelligence News about Threat Actors*"

Dumme Maschinen

AI / KI / NLP Forschung benötigt saubere Corpora pro Wissensdomäne

Existierende (Open-Source) NLP Software liefert meist grundlegende Modelle, die auf Internet und News Daten basieren.

Für echte, “Cyber”-spezifische Textklassifizierung und Labeling brauchen wir Cyber Corpora, die öffentlich zugänglich und für akademische Zwecke verifizierbar sind.



TI Report Quellen

Es gibt viele öffentliche (TLP:CLEAR) CTI Report Quellen mit sehr unterschiedlichem Grad an maschinenlesbarem Zugriff und Kontextinformationen.

Informationen werden meist als Marketing Publikation bereitgestellt und können später durch M&A etc. de-publiziert oder unzugänglich werden.

Das führt zu erheblichem Wissensverlust.

ORKL Architektur

Library Manager

Kuratiert einen Dateibasierten Corpus und fügt neue Reports hinzu

ORKL API

Zugriff auf alle Daten in der ORKL Datenbank

ORKL Frontend

Zugriff auf die wichtigsten Funktionen der API für Endnutzer

ORKL Cyber Threat Intelligence Library

Eine öffentliche Bibliothek für CTI Reports ohne Überziehungsgebühren :)

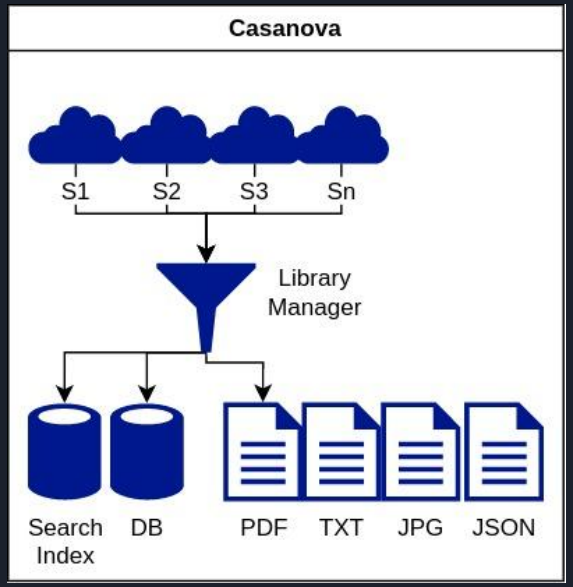
Das Casanova Quadrumvirate

Für jeden hinzugefügten Report legt der Library Manager 4 Dateien an:

- PDF → Originaldatei (unverändert)
- TXT → Textrepräsentation
- JSON → Metadaten + Text als JSON obj
- JPEG → Bild der ersten Seite

Wenn ein Report in mehreren Quellen enthalten ist, wird er nur einmal gespeichert, Quellen Informationen werden aber erhalten.

Die Metadaten aus den verschiedenen Quellen werden gebündelt im Library Entry gespeichert.



Threat Actor Profile

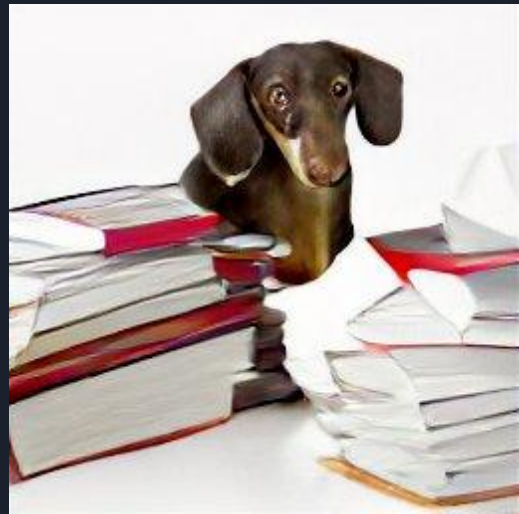
Der Library Manager pflegt auch Threat Actor Profile, wenn die Quelle dies anbietet.

Für Querverweise in der Datenbank sind vor allem Synonyme/Aliase und gebräuchliche Namen der Threat Actor Profile relevant. Das gleiche gilt für verwendete Tools.

Die Quelle der Informationen bleibt in jedem Schritt erhalten.

Die Profile können als Suchvektoren für den Suchindex verwendet und zu den Reports gemappt werden.

Die Qualität der einzelnen Namen kann mit Frequency Analysen bestimmt werden (z.Bsp. TF-IDF)



Malpedia	https://malpedia.caad.fkie.fraunhofer.de
Alienvault OTX	https://otx.alienvault.com
ETDA Threat Actor Library	https://apt.etda.or.th
CyberMonitor	https://github.com/CyberMonitor/APT_CyberCriminal_Campagin_Collections
APTNotes	https://github.com/aptnotes
SecureWorks	https://www.secureworks.com/research/threat-profiles
MITRE ATT&CK® Data	https://github.com/mitre-attack/attack-stix-data
APT Groups & Operations	https://apt.threattracking.com
MISP Galaxies	https://github.com/MISP/misp-galaxy
VX Underground	https://www.vx-underground.org

Eine Liste der aktuellen Quellen

Community Partizipation

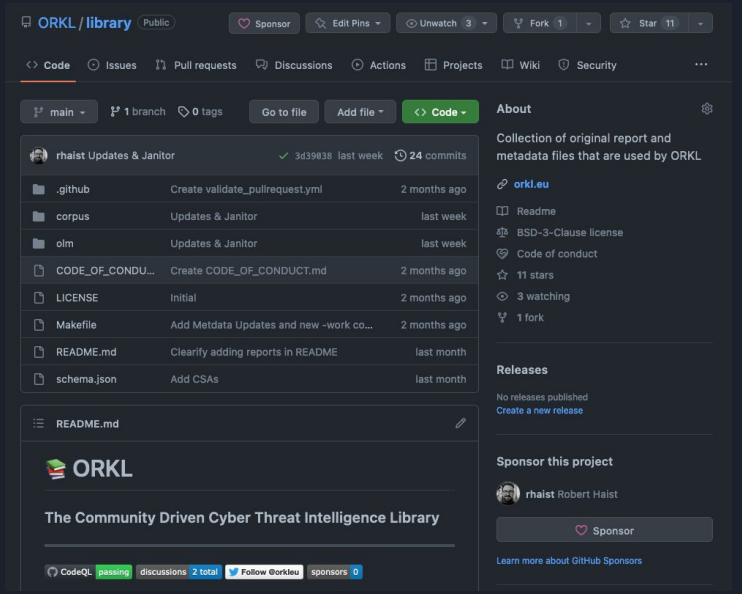
Git-basiertes crowdsourcing von Metadaten Updates

OLM Metadaten Hilfsprogramm (Clippy für ORKL)

Hinzufügen von neuen Reports

Automatisches Updaten des CDN und der ORKL DB

https://github.com/ORKL/library



ORKL API

Vollzugriff auf den aktuellen Stand der Bibliotheksdatenbank

Volltextsuche

Links zu den Dateien im CDN

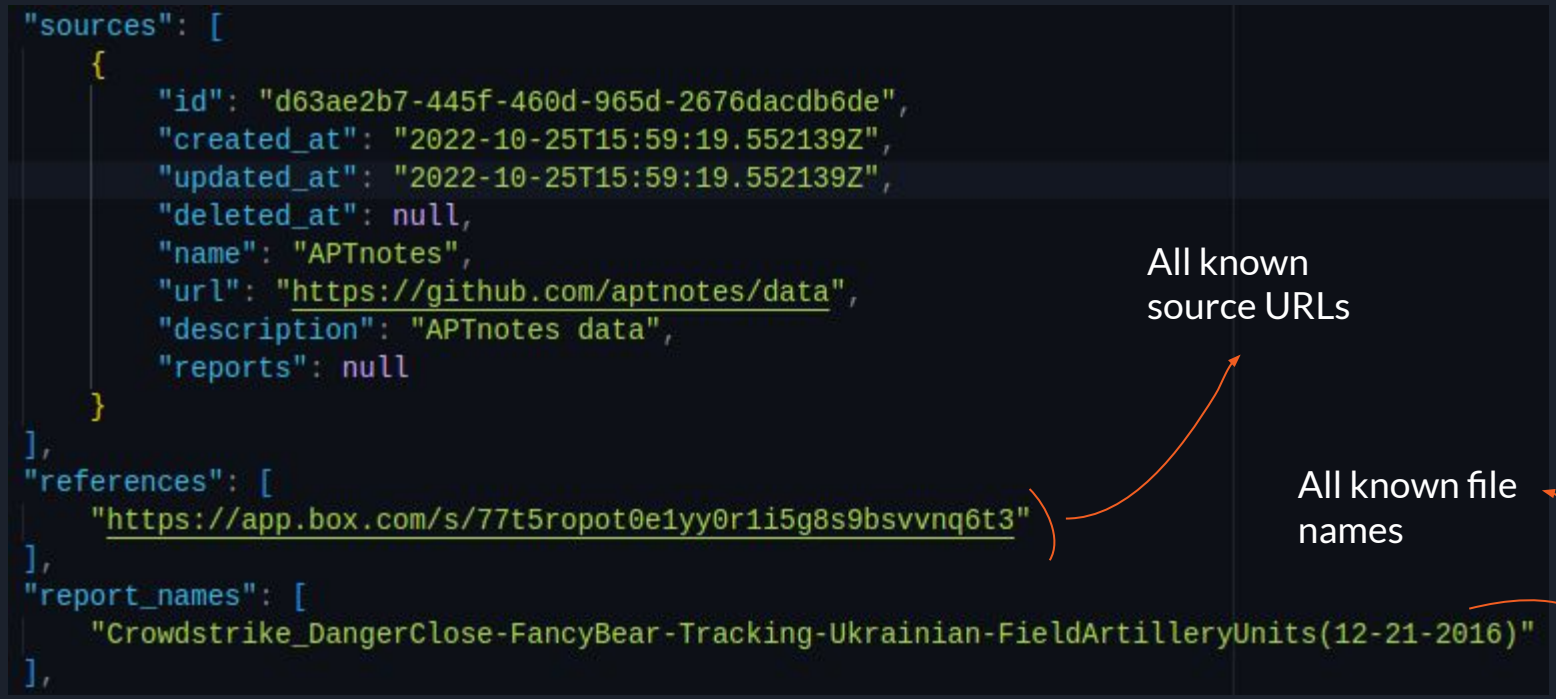
Quellen Informationen

Threat Actor Profile

```
{
    "data": {
        "id": "67a2c542-0506-4eb8-8afd-20d0e757bf0c",
        "created_at": "2022-10-25T16:48:25.06851Z",
        "updated_at": "2022-10-28T13:16:04.976132Z",
        "deleted_at": null,
        "sha1_hash": "860387572ad036bfde33775ee89e7d92fa5d0aae",
        "title": "Danger Close: Fancy Bear Tracking of Ukrainian Field Artillery Units",
        "authors": "Crowdstrike",
        "file_creation_date": "2017-07-27T03:00:51Z",
        "file_modification_date": "0001-01-01T00:00:00Z",
        "file_size": 262427,
        "plain_text": "Danger Close: Fancy Bear Tracking of Ukrainian Field Artillery Units\n\n\n<SNIP>"
```


snip for readability

Example Library Entry: PDF/Source based Metadata



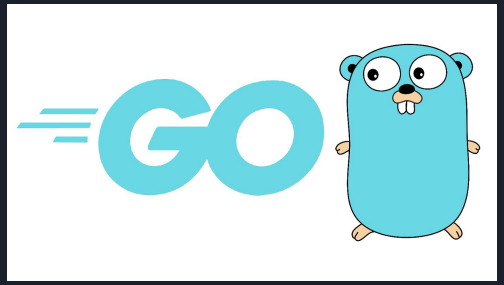
Example Entry: Continued

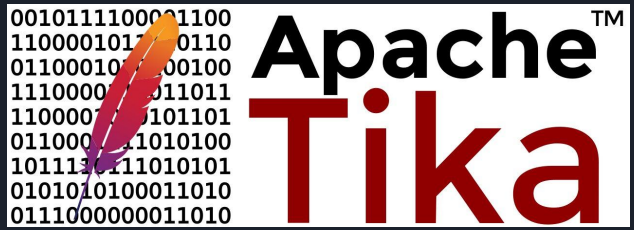
Combine Source and MainName to reference the Object throughout the App

Example Entry: Threat Actor association

Example Entry: Threat Actor association

Gebaut mit Open Source ❤️





Roadmap

UI	Crowdsourced Librarian	NLP
Das Frontend soll einen besseren Zugriff von allen Plattformen (Mobile & Desktop) ermöglichen.	ORKL erlaubt schon, dass die Community Metadaten von Reports pflegt. Leider gab es keinen Community PR bis jetzt	Training eines Cyber Security fokussierten NLP models für ein Open Source framework wie spaCy

Community Unterstützung

Mehr Quellen

- Zusätzliche, maschinenlesbare Quellen für Reports und Threat Actor Profile

UI

- Frontend Magier
- Logo

Metadaten

Es gibt jede Menge Metadaten zu korrigieren

https://orkl.eu/contribution

orkl.eu

Viel Spass beim Stöbern

Twitter Updates  @orkleu

Hosted by 

Kontakt

@RobertHaist

@rhaist

Referenzen

Folie 3: Bild via www.vpnsrus.com

Folie 4: Various front pages of sample CTI reports - copyright remains with the original authors

Folie 7: https://commons.wikimedia.org/wiki/File:Hello_my_name_is_sticker.svg

Folie 10: https://commons.wikimedia.org/wiki/File:Rijks_Museum_Library.jpg

Folie 15: Various Open Source Project logos - copyright remains with the original creators