AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

ADVANCED OSINT ANALYSIS FOR NIS AUTHORITIES, CSIRT TEAMS AND ORGANISATIONS

IKT Sicherheitskonferenz des AbwA

TARANIS_AI
https://taranis.ai/

Florian Skopik and **Benjamin Akhras**
Center for Digital Safety and Security
AIT Austrian Institute of Technology

Linz, October 3rd, 2023

WHAT IS CYBER SECURITY OSINT?

- Open-source intelligence (OSINT) is the collection and analysis of data gathered from open sources to produce actionable intelligence.
- Technical Cyber Threat Intelligence (CTI) to configure detection systems:
  - Indicators to put into SIEMs
  - Domains to block in name-servers or proxies
  - Execution patterns to block in EDRs
- But also “softer CTI”:
  - News about threat actors
  - New (features of) security products
  - News about breaches, incidents, campaigns
  - News about vulnerabilities, patches, mitigations, counter-measures, exploitation, post-exploitation, …
  - Policy news: political/diplomatic initiatives, new EU policy documents, GDPR-related lawsuits
  - Updates on security standards (ISO, BSI, ANSI, CIS, OWASP, …)
  - Mergers, acquisitions, failures, … or other company news

Baseline today at CERT.at: approx 250 sources, up to 500 articles per day, more after longer weekends or large-scale events

WHAT IS OSINT GOOD FOR?

- Gather public information on potential security threats, vulnerabilities, trends, attacker TTPs, new risks etc. to maintain situational awareness and take early counter actions.
- Input for products
  - Advisories
  - Summaries (daily, weekly)
  - Situational picture reports, white papers, fact-sheets
- Awareness / Preparedness
  - Consulting / Answering calls for help
  - Media inquiries
  - “Boss/CEO/Politician asking questions”
  - Trigger for proactive activities
- Challenge: Number of OSINT sources is high and the number of news items massive
  - Grasp quickly what’s relevant and omit the rest
  - Filter repetitive content
  - The workflow is actually pretty similar to a journalists work

CEF PROJECT AWAKE

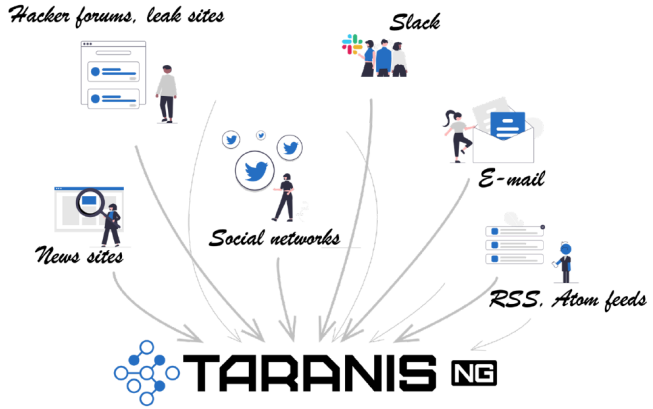
Co-financed by the Connecting Europe Facility of the European Union

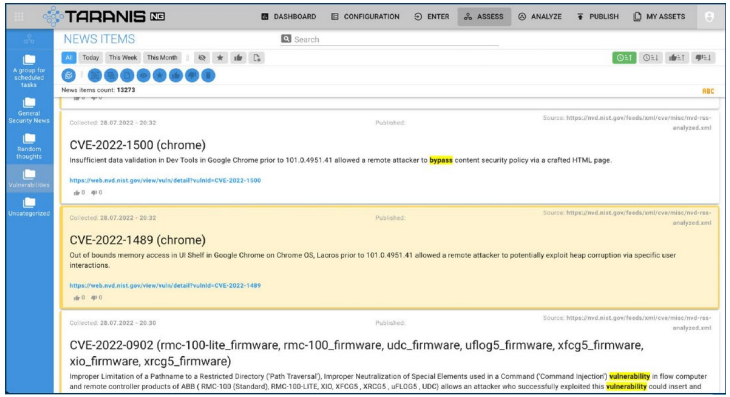
- CEF funded project:
  - AIT (coordinator), CERT.at, Ministry of the Interior (NIS-CA), Federal Chancellery
- Main Goal
  - Sharing Situational Awareness between the NIS actors in Austria (CSIRT, NIS-CA(s), LEA, MFA)
  - By collecting, aggregating, filtering, translating and sharing situational information
- Focus on
  - Unstructured data
  - Freeform text
  - For human consumption
  - Focus is **not** on technical or operational data, but on high-level, strategic information

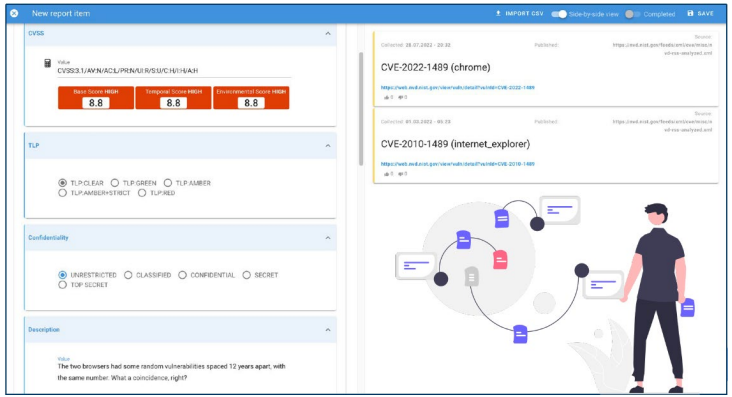
https://github.com/SK-CERT/Taranis-NG

TARANIS NG

- Ingest raw unstructured data from various sources
- Use human knowledge to identify relevant information
- Compile reports and distribute advisories and reports to the community
- **→ A great tool and an awesome basis for further extensions using modern NLP and machine learning …**



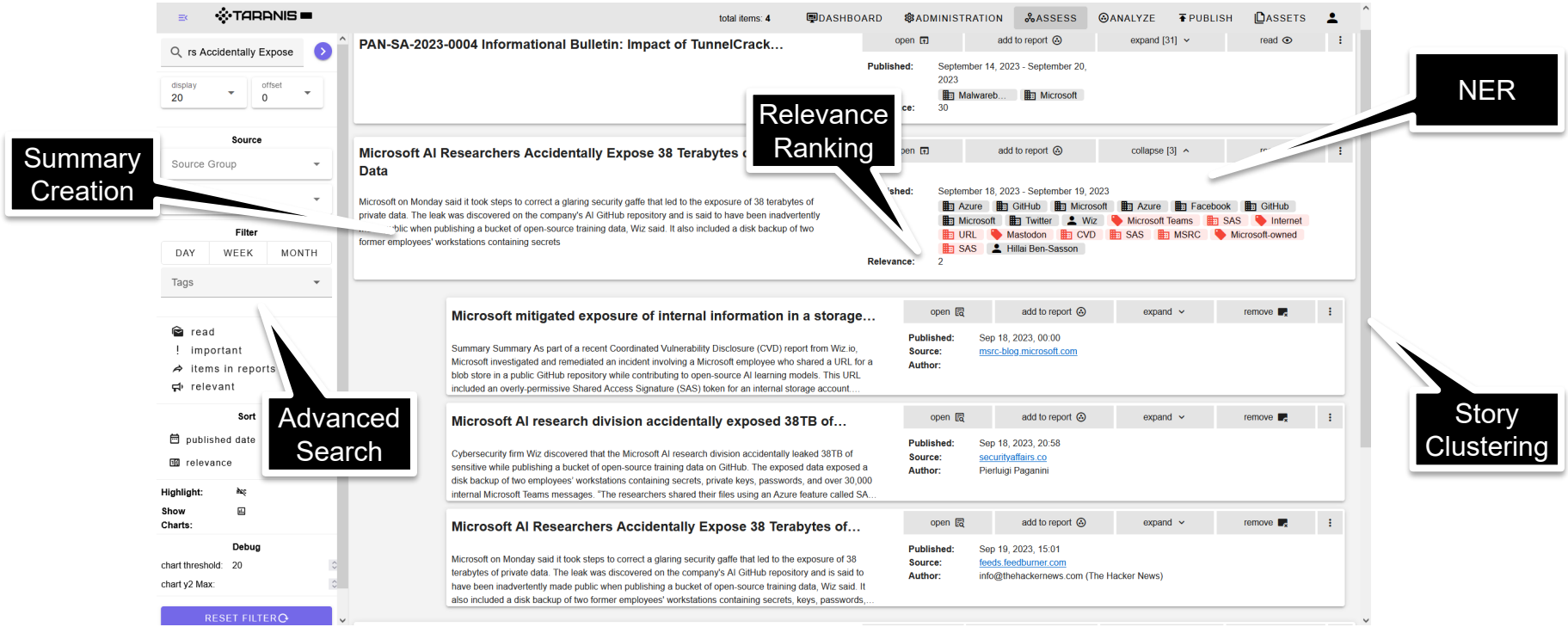




APPLYING THE POWER OF NLP

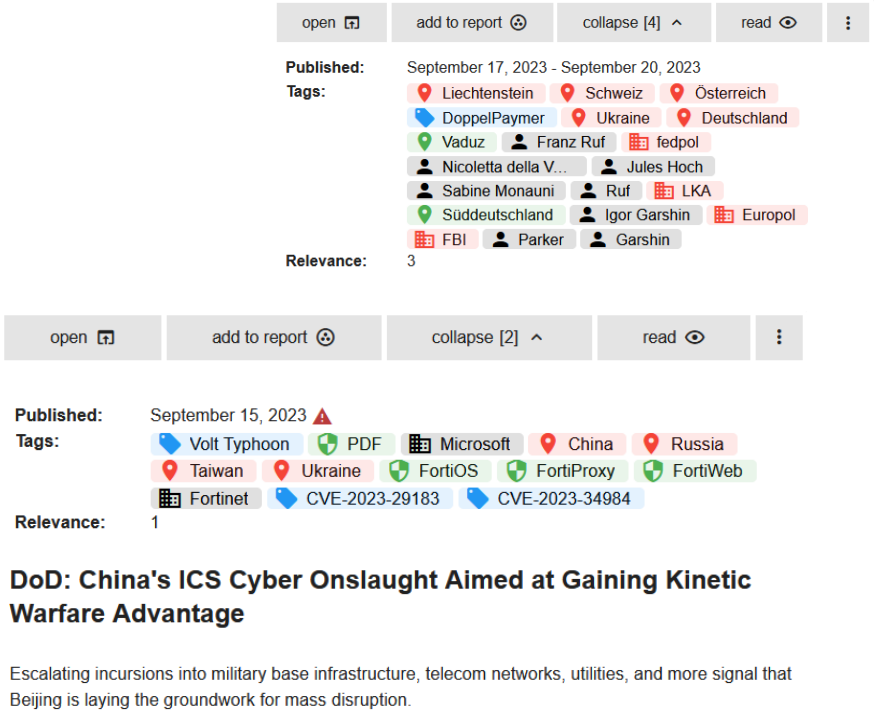
- User Story 1: What was going on in the last 24 hours? („Hot Topics Clustering“)
- User Story 2: What do we know about entity X? (e.g., a vulnerability, malware, company, product)
- User Story 3: I‘ve read an interesting article. What related news items do exist?
- User Story 4: Which news items are recommended (collaboratively and AI-assisted)?
- User Story 5: I‘d like to build a sharing set for my partners. How to sum up my findings?

NOVEL FEATURES AND DEVELOPMENTS



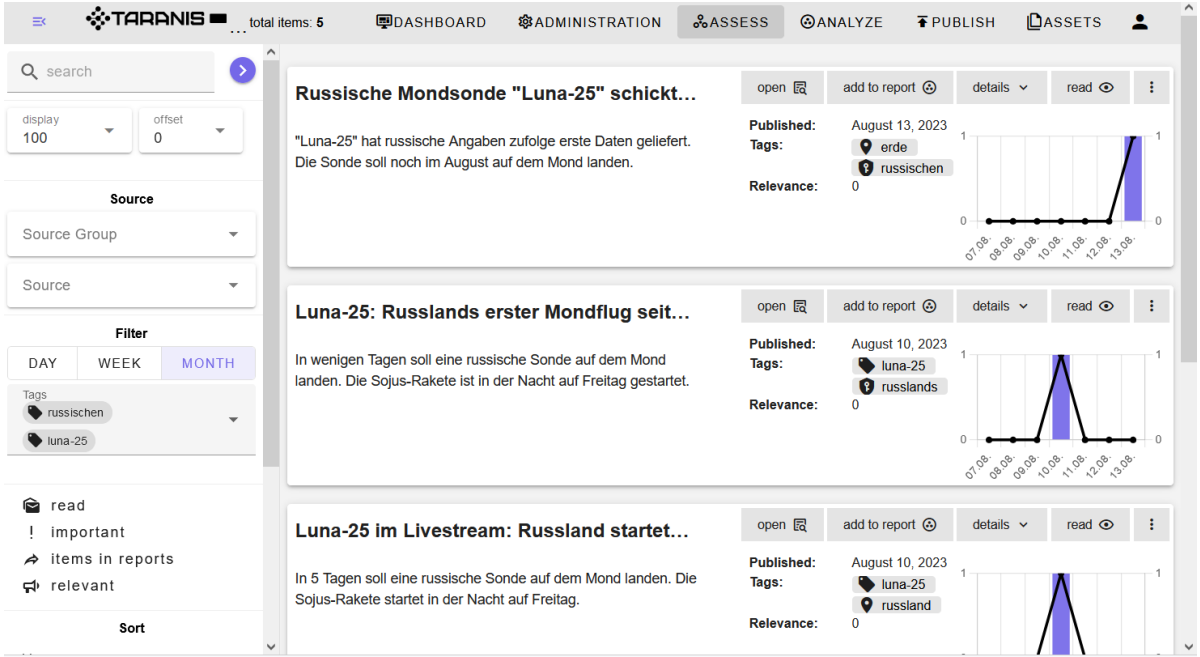
FEATURES (1/5): NAMED ENTITY RECOGNITION

- Named-entity recognition (NER) seeks to locate and classify named entities mentioned in unstructured text into pre-defined categories such as person names, organizations, locations, etc.
- Mix of pre-defined lists (e.g., countries), custom regex (e.g., CVEs), and trained language model on German/English standard text
- Custom extensions to recognize IT products, vendors, APT groups etc.
- *Future Extension:*
  - Additional use of domain-specific word lists



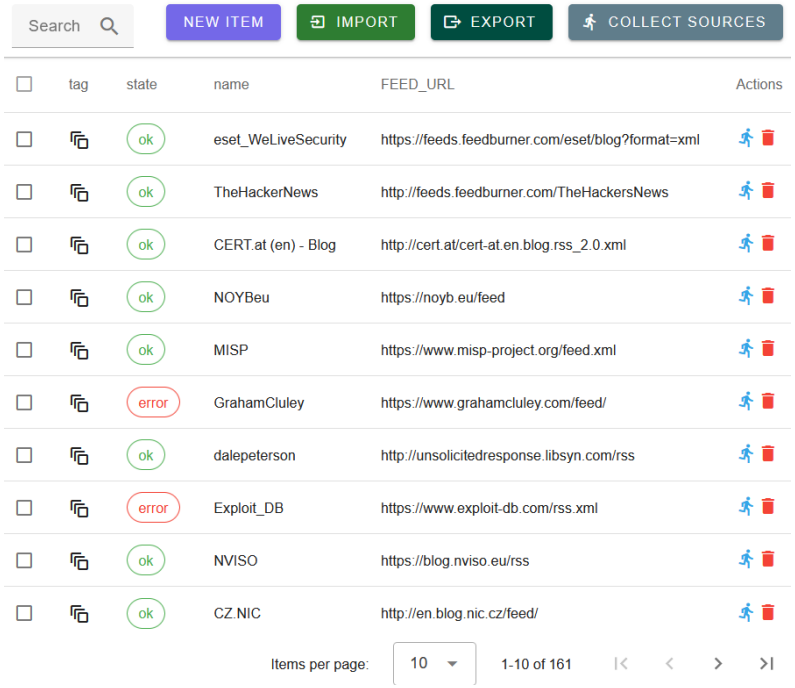
FEATURES (2/5): ADVANCED SEARCH & FILTERING

- Tags from NER can be used to filter and cluster items belonging to the same topic
- Additional free-text search
- Further filter and sort parameters (read/unread, relevance score etc.)
- *Future extension:*
  - Collaborative search and filtering
  - Sharing of filters



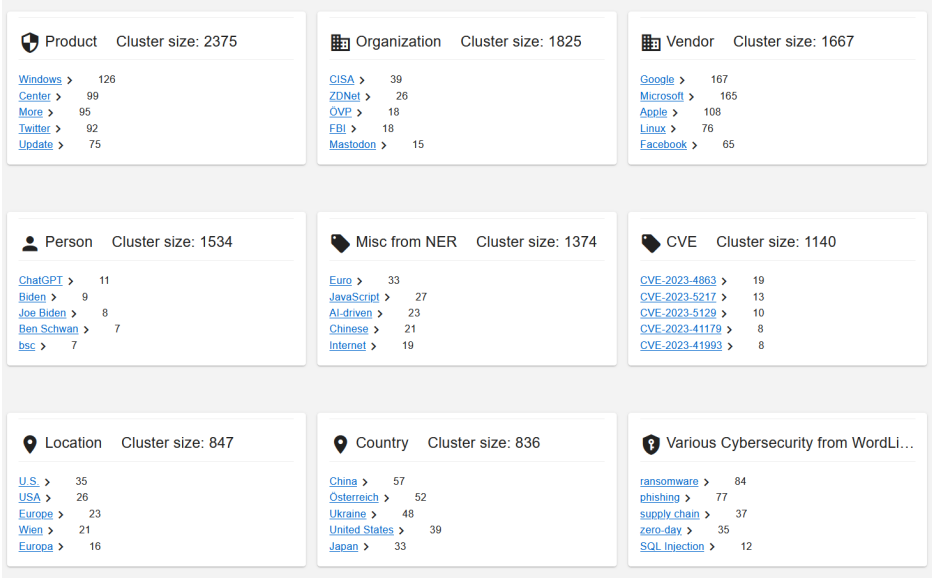
FEATURES (3/5): RELEVANCE RANKING

- Relevance Ranking helps to identify “interesting items” based on general importance and personal preferences
  - Upvotes and downvotes from collaborators
  - Shared articles
  - Related news items
- *Future Extensions:*
  - Feedback Loop: Learn properties of often up-/downvotes items.
    - What properties do “good” articles have in common?
    - Which sources deliver such items?



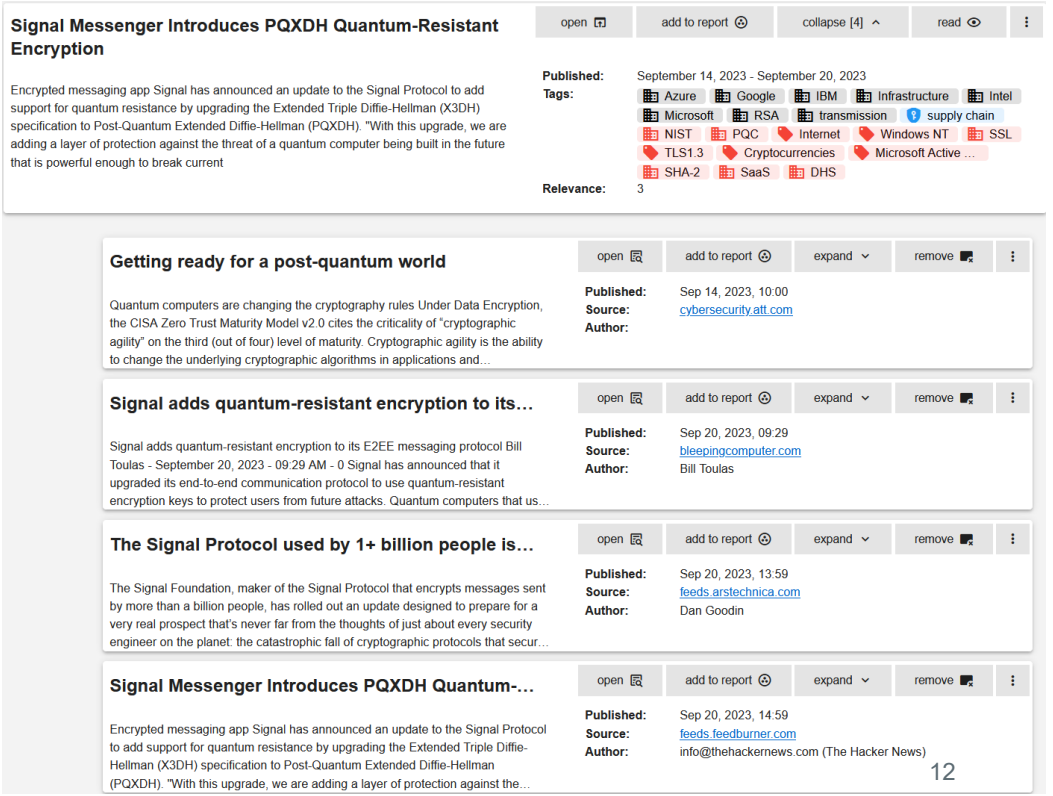
FEATURES (4/5): SUMMARY & REPORT CREATION

- Summaries help to condense lengthy texts to their essential parts for quick decisions on their relevance
  - Summary of lengthy news items to quickly grasp its content
  - Summary of stories based on its collection of articles
- *Future Extensions*
  - Summary of Sharing Sets for reports
    - AI-assisted Pre-filling of report fields
  - Tuning of summaries regarding appropriate length, wording and content



FEATURES (5/5): TOPIC & STORY CLUSTERING

- News Items are usually delivered not just by one, but multiple sources at approximately the same time with mostly similar content
- Cluster items and create “meta item” that summarizes important content
  - Decrease human effort needed to ingest all news items!
  - Visualize development of a story over time
- Show development of “top topics” over time



LIVE DEMO

CONCLUSION AND OUTLOOK

- AWAKE: several NLP extensions that improve taranis-ng to increase the efficiency of OSINT analysts by clustering, categorizing and filtering items
- Continuous development secured!
  - EDF and DEP projects until 2027
  - Collaboration with CERT.sk -> upstream of changes!
  - Opportunity to contribute: https://taranis.ai/
- Planned extensions:
  - Allow the collaboration across multiple instances, e.g. CERT, ministries (future extension)
  - Matching of news items to predefined asset lists (CPEs) to determine risks
  - Improvement of NLP magic
- Future application cases
  - Different from SOC cases
  - Coordination of NIS-relevant knowledge to specific topics between CERT and BM*

AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

THANK YOU!

TARANIS_AI
https://taranis.ai/

Please contact:

Florian Skopik

florian.skopik@ait.ac.at

October, 3rd 2023