**Task.** This task will focus on the definition of methods and indicators for the risk assessment of new data sources. Risk level is referred to a different aspects strictly connected: on one hand, the risk level is due to the involvement of source in illegal trafficking activities; on the other hand, the possibility that criminal groups that own or manage sources monitored by LEAs could recognise anomalous activities due to the crawling process and block them, compromising the investigations. Taking into account that risk level of a source can increase with its depth in the Web, such methods will then consider the position of the source (in Surface Web, Deep Web or Dark Nets), metadata that can be extracted in order to define which policies should be enabled to perform the crawling and with which degree of anonymity. Finally, websites appearance, their update frequency and refresh in data will be taken into account during the assessment stage. The outcome of this task consists in a set of guidelines, methods and principles to establish the risk level of sources in Surface Web, Deep Web and Dark Nets, and the related policies to be applied for ensuring safe and secure crawling.

**RQs.**

MRQ.what guidelines, methods, and principles exist to establish cyberthreat level of online sources?

SRQ1. what online depth levels are more assessed?

SRQ2. what degrees of anonymity exist for web-crawling?

SRQ3. what policies exist to vary the degrees of anonymity?

SRQ4. what website features are most indicative of cyberthreats?

SRQ5. what risk assessment techniques exist?

SRQ6. what data sources exist and how are they classified?

**Method.**

MSLR

**Search Query.**

(Guideline\* V Method\* V Principle\* V assessment) AND (Cyber AND (Threat V Crime)) AND (Online V Web V Site V Website V Source) AND (Surface V Shallow V Deep V Dark) AND (risk OR issue OR threat)

**Sources.**

ACM DL

IEEEXplore

Wiley Interscience

Scopus

Bibsonomy

**Inclusion/Exclusion Criteria.**

|  |  |
| --- | --- |
| INCLUSION | EXCLUSION |
| Workshop paper needs to be co-located with CORE A or A\* | Paper is more than 5 y.o. \*and\* received 0 citations. |
| Paper actually matches (Guideline\* V Method\* V Principle\*) AND (Cyber AND (Threat V Crime)) AND (Online V Web V Site V Website V Source) | Paper is not published or not available online as a pre-print |
| Paper contains some form of evaluation | Paper discusses a minor result in a lesser venue (< CORE Rank A) |

**Query Results.**

RschLit. 189 results

GreyLit. 166 results

Excl. 29 Tot.