INFINITY

IMMERSE. INTERACT. INVESTIGATE.

Technical Brief

Opportunities in AI Patterns

DISSEMINATION LEVEL PUBLIC

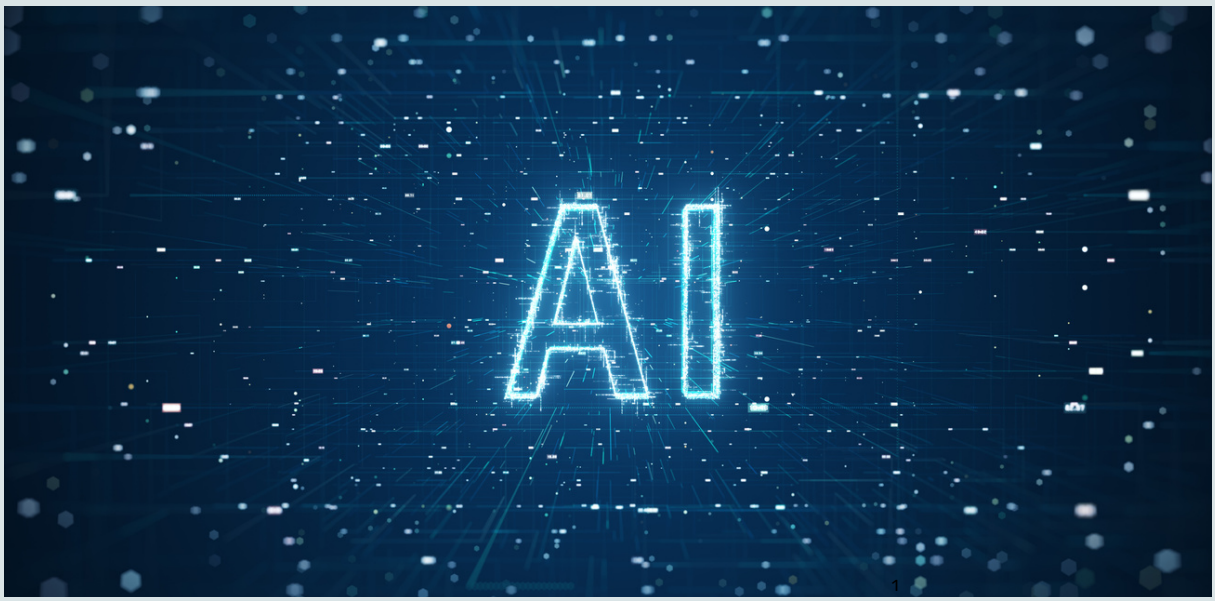
PARTNER

ENG

AUTHOR

Marialuna De Tommaso
Roberto Acquaviva
Silvio Sorace

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 883293. The content of this document represents the view of the authors only and is their sole responsibility. The European Commission does not accept any responsibilty for any use that may be made of the information it contains.



1. Crime Investigation Opportunities in AI Patterns[1]

A crime investigation is a collection of events/actions, resources of interest, relevant information, items, entities and relations to describe the dynamics that led to the realisation of a crime.

Such heterogeneous data, managed by INFINITY, are stored by the Knowledge Base component and represented as a Knowledge Graph (KG). The Knowledge Base (and, consequently, the KG) is constantly enriched with new data coming from tools and services integrated into the platform as well as data created by the investigators.

In this context, a pattern is a meaningful set of nodes and relations, and it is represented by a graph.

Having defined and structured the Knowledge Base as a graph, patterns are useful to recognise and describe meaningful subgraphs. They are also important to simulate potential scenarios for the lack of information in the crime storytelling, enabling the possibility to look for missing entities in the reconstruction of the salient facts of the case.

Patterns are modelled as graphs and pattern recognition amounts to finding a correspondence between the nodes of different graphs and their relations.

The process of information storage is very complex due to the large amount of heterogeneous data coming from analysis modules and users' operations. The provided information is rich (in features and relations), making not immediate their semantic representation. It is necessary to understand the basic definition of the INFINITY information model, in order to effectively define a good knowledge organisation in support of information management.

Information becomes knowledge only when semantically contextualized. This implies that a domain modelling stage is required to achieve a uniform interpretation of information and its correspondent representation into knowledge. To do this, the common approach is the development of a shared operational ontology:[I] by using an ontology, a universal semantic can be provided to foster interoperability and information exchange.

[I] An ontology is used to model a domain of knowledge. It consists of a set of concepts and relationships. These relationships go over than the simple parent-child relationship.

2. INFINITY Ontology-driven approach

In INFINITY, ontologies are strongly involved in the process of representation of entities and relationships that deal with the Intelligence activities. Starting from the study of the state-of-the-art ontologies, the INFINITY operational ontology has been designed as a solid ground truth for the entire knowledge management process. In this section, we first review the state-of-the-art ontologies and then describe the analysis process used to derive the INFINITY operational ontology.

2.1 State of the art of ontologies for terrorist-related and cybercrime activities

In the past, some general ontologies have been proposed in order to structure a unique shared model to represent knowledge. However, these approaches were considered in many cases limiting, because of the need to represent particular details of events and information according to the application context and of users that have to use it.

This is also the case of the INFINITY operational ontology. The study of existing terrorist-related and cybercrime ontologies revealed the need of a specific ad-hoc ontology, in order to cover all the use case applications and to provide a heterogeneous integration of information extracted by modules.

A method to recognize and evaluate terrorist threats based on an ontology data model is proposed in Najgebauer et al 2008.[2] A semantic graph is used to represent knowledge in terms of facts and events: through an analysis process, it is possible to acquire indirect associations, which allow pinpointing new knowledge, in term of relationships, indicating possible threats. Different approaches of threats extraction have been defined, all of them based on a rule-based inference over a graph knowledge structure.

In order to represent all stored data and relationships in a semantic graph, an ontology is used (Figure 1), having the following taxonomy[II] structure: a Terrorist organization, a Target object and a Registered event are the main components associated with a Terrorist threat.

Figure 1: Terrorist ontology – Taxonomy structure

[II] A taxonomy, as well as an ontology, is used to model a domain of knowledge, and consists of a set of concepts and relationships. The difference is that a taxonomy uses a parent-child relationship in its classification.

System (EWS), a simulation-based diagnostic support tool able to collect relevant information to terrorism threat estimation and intelligence data analysis.

Turner et al. 2011[3] present a foundation ontology, aiming to represent adversary groups and their intentions, classify their weapons and attack types, and represent the relationship between the outcomes of an attack and the various recognized intentions of the adversary group. The Adversary-Intent-Target (AIT) model is compliant with the Basic Formal Ontology (BFO),[4] which divides concepts into material entities and their parts, the qualities (handled by material entities) and processes (where material entities are involved). According to the previous definition, BFO is extended with the following AIT definition of "Terrorist Attack": A terrorist attack occurs when an adversary, with intent and capability, uses a weapon against a target"

Thus, a TerroristAttack is an Event with a minimum of one Target, using a minimum of one Weapon, and involving a minimum of one AdversaryOrganization as an agent that has both the Capability and Intent.

A portion of the ontology that relates the TerroristAttack type with the kind of Intent held by various AdversaryOrganization types through the type of Outcome, is shown in Figure 2.

This model of knowledge allows intelligence analysis to reason about the key elements of a terrorist attack:

- Who (existing terrorist organization and their characteristics);
- What (the kind of weapon used for specific attacks and the related outcome);
- Where (the possible target);
- Why (the intentions that lead to choose a particular type of attack).

Figure 2: AIT Model – Terrorist attack representation

An ontology-based automatic event extraction system for the prediction of which terrorist group is most likely to be responsible for an event is proposed in Inyaem et al. 2010.[5] The goal of this process is to extract all the relevant instances of a terrorism event, like victims, date, places, time and tactics. The results are used as input into the prediction process of the terrorist groups involved in a terrorism event (Figure 3).

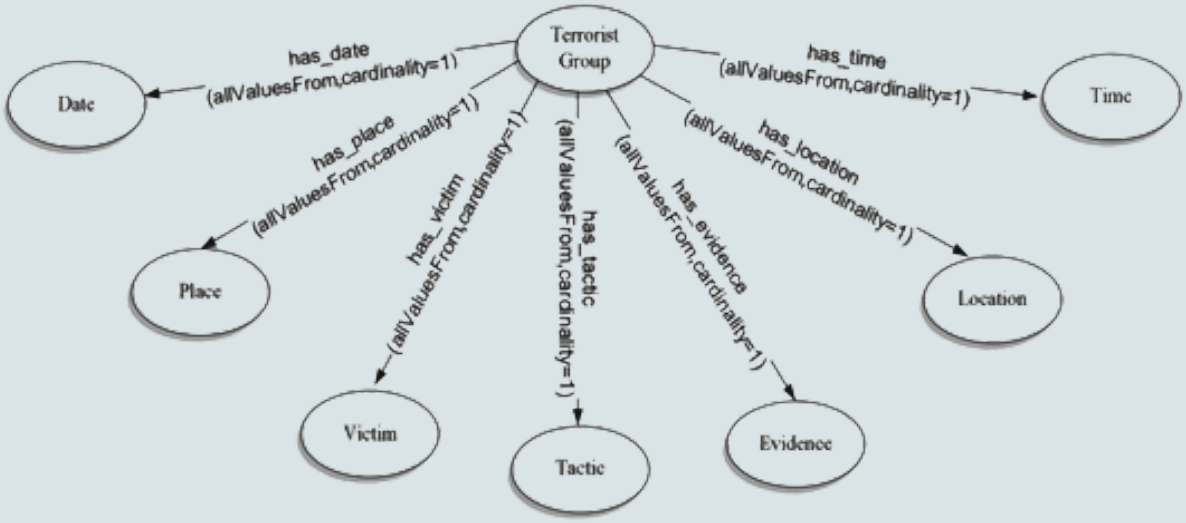


Figure 3: A terrorism ontology – N-ary relations

All the mentioned studies involve the use of "Terrorist Attack" entity: this is a strong starting point to create a more detailed ontology where Terrorist Attack can be considered an "event". In recent years, the increase in cybercrimes (and its related negative impacts on the lives of individuals, organizations, and governments) has led to the consideration that a better understanding of cybercrime is a necessary condition to be able to provide a meaningful representation of events of interest to Intelligence activities.

The term "Cybercrime" is generally used to cover a wide variety of illegal crimes or what is considered illicit among traditional crimes that are contingent on the use of technology, publication of illegal content on the internet, and crimes that occur within technological forums. A comprehensive cybercrime classification ontology that incorporates multiple perspectives (e.g., Attack, Event, Attacker, Impact, Objective, Victim, Target, Complainant, Vulnerability, Tool and Technique, Location and Offence) is presented in Donalds et al.[6] These perspectives are included in the ontology as they would provide pertinent information that would be beneficial for police organizations to classify day-to-day cybercrimes, identify trends and patterns and issue releases for public education. The proposed ontology integrates and extends prior cybercrime classification schemes by incorporating previously proposed perspectives as well as adding new perspectives, respectively. The study offers: i) a comprehensive and multi-perspective cybercrime classification and analysis model that involves a set of relevant concepts and direct relationships; ii) a flexible artifact that is better able to classify current and future cybercrime attacks; and iii) a formal Cybercrime Classification Ontology that could allow cybercrime agents (human or artificial) to share knowledge.

Also, the work described in Barn[7] contributes a semi-formal approach to the development of a taxonomy for cybercrime and offers the conceptual language and accompanying constraints with which to describe cybercrime examples. The approach uses the ontology development platform, Protégé and the Unified Modelling Language (UML) to present an initial taxonomy for cybercrime, whose representation of a cyberattack is shown in the picture below.

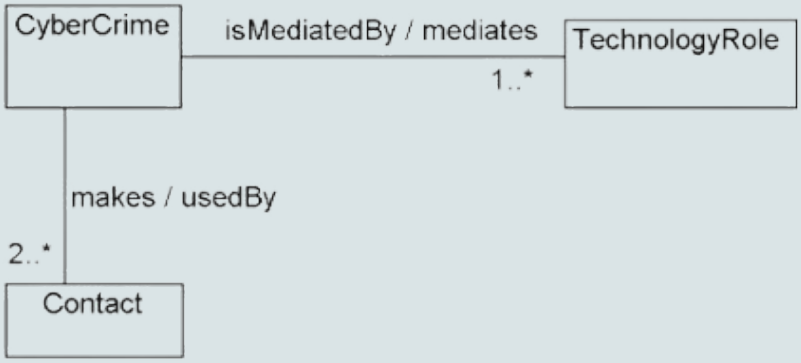


Figure 4: Representation of a Cybercrime attack

The main limitation of all the ontologies presented above, preventing their direct adoption for INFINITY, is their lack of support for the various information sources that are supported in the project as well as the use-case complexities that need to be handled by the system. However, the above ontologies, constituted a helpful reference in the definition of the current INFINITY operational ontology. An important contribution comes from security events ontologies, mainly in the representation of the Event entity as an aggregation of relevant information representing time, space and agent. Looking at the AIT model, we have tried to represent accurately the relationships that exist between the main entities, also defining a basic support to reasoning mechanisms.

2.2 INFINITY operational ontology

The INFINITY operational ontology is a specific ontology that has been defined to easily model information related to security events to be stored into the Knowledge Graph. The ontology can be considered as a model that depicts the KG data structure, where concepts rule the types of information that can be stored into the KG, e.g. People, Objects, Groups (intended as set of people), Events, Resources (i.e., multimedia content like Texts, Images, Videos, Audio files and e-mails) and Web Sources.

	INFINITY KB operational ontology	
	Name	**Description**
Concept	Resource	A multimedia content of interest
	Event	An activity that took place
	WebSource	e.g., a website, a forum, a blog, etc.
	Person	An individual
	Organisation	People organized into groups
	Location	A site or position.
	SocialUser	The identity of a person on the web
	Object	An object that participates in an activity

Due to the organisation of knowledge into entities connected with each other by relationships, the knowledge representation process takes place through a dynamic aggregation of relevant entities participating in a criminal activity: to this end, a set of meaningful relationships to provide the right representation of knowledge is required. The KG operational ontology provides all the properties characterizing this process: entities can relate to each other through a set of object properties representing the domain relationships allowed between the involved entities.

It is very important to respect the structure of relationships expected between different couples of entities, to optimize the results achieved by the pattern recognition. Indeed, a very important benefit, in the exploitation of a graph to store the knowledge, is to know in advance its structure: in the same way, the relationships that come in/come out as an entity and which entities will be at the extremities of a specific relation are well-known. Ontologies have been proposed not only to enhance the knowledge representation and storage, but also to improve reasoning capabilities with rule-based mechanisms based on entities handled in KG operational ontologies and their properties. These mechanisms allow to identify sequential patterns of simple events and translate them into more complex illegal activities. Moreover, data mining techniques can also be applied to identify new patterns of events and then generating new rules for the inference of additional complex activities. These techniques have been applied in INFINITY to provide LEAs with support to the discovery of insights and decision-making activities. The main goal of the AI driven pattern recognition is to detect and recognize new information that might be of interest for the INFINITY domain.

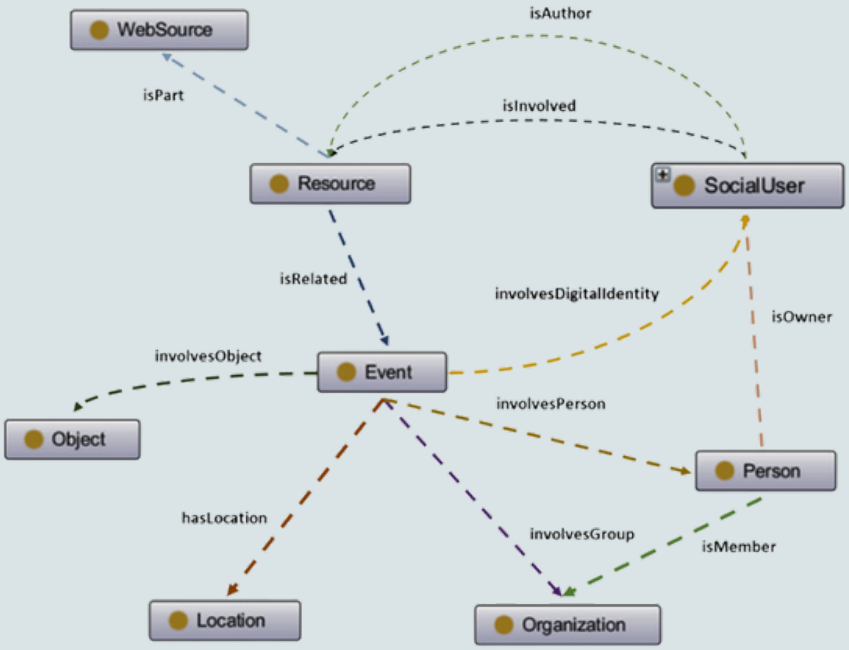


Figure 5: INFINITY operational ontology – structure

3. Knowledge Patterns

As mentioned in the previous paragraph, the representation of knowledge in specific patterns, relevant to the identification of suspicious event and illegal activities, is the basis of reasoning mechanisms.

A Knowledge Pattern (KP) is the topological[III] definition of a part of a knowledge graph (a sub-graph) that can be relevant to the domain of application. Generally, a sub-graph contains more information than the sum of single entities that compose it, since it also involves the relationships (with types and metadata) existing among the involved entities. This makes KPs suitable to represent complex information that exist among one or more entities and that would not be represented with the same level of accuracy by considering each single entity at time.

The approach adopted for carrying out activities of the AI-driven pattern recognition is based on the definition of strategies to find out complex illegal activities through known KPs, as well as the identification/discovery of new patterns.

Information coming from modules, responsible for the extraction of data from multimedia resources, can be represented by basic "low-level" knowledge patterns.

KPs with increasing complexity can be identified and composed gathering low-level patterns: this process allows to discover high-level knowledge starting from low-level information about events, people, groups, organizations, objects, resources and their relationships. Figure 6 shows a concrete example given by the aggregation of PSu (Person – Social user) and SuR (Social user - Resource) patterns, resulting in the PSuR (Person – Social user - Resource) complex pattern. The PSuR pattern collects information about the account held by the person on the web and the resources created by that account.

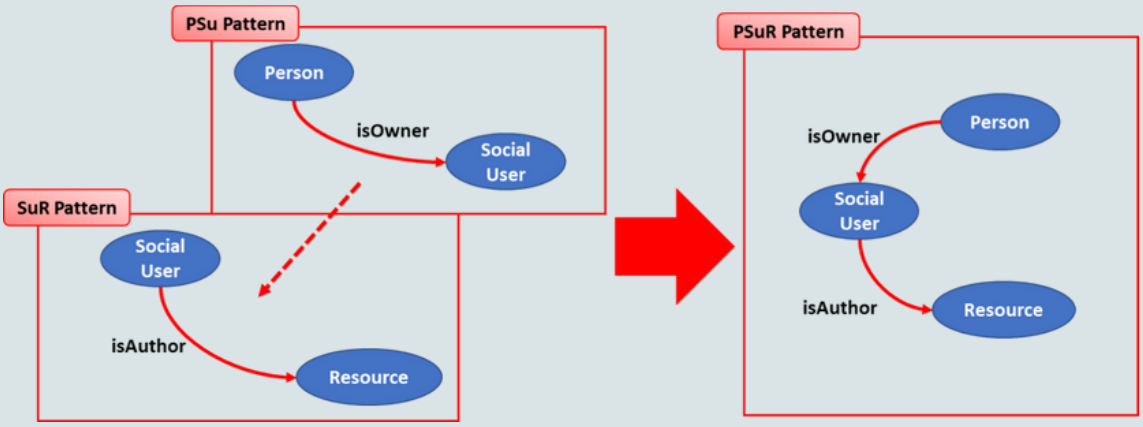


Figure 6: Patterns aggregation

Also, KPs have a well-known topology: due to the organisation of knowledge into entities connected with each other by relationships (described in the previous paragraph) it is also possible to identify missing pieces of knowledge and incomplete KPs starting from expected topologies. An incomplete pattern can be useful to suggest users which kind of missing information to look for in order to complete a pattern. In this way, the process can alert users about potential suspicious events that can occur and guide in tracking illegal activities.

[III] A topological space is a set endowed with a structure, called a topology. It does not matter if the relationships are "parent-child (like in the taxonomies) or more complex (like in the ontologies).

4. Pattern Recognition

The INFINITY Knowledge Graph is constantly enriched with new data: results coming from analysis of resources can be merged with available previous knowledge in order to detect and recognize new information that might be of interest for crime domain. A very important benefit, of defining a model to organize knowledge to be stored into a KG, is to know in advance the relationships that come in/come out as an entity and which entities will be at the extremities of a specific relation. Based on these premises, the approach adopted allows to identify missing pieces of information through a knowledge enrichment process. The aim is to suggest new knowledge, in terms of patterns composed of missing entities and relationships between them, to support the resolution of the investigation through the reconstruction of criminal activities. The most important pattern, relevant to the INFINITY domain, is the Similarity pattern which will be used in the process of knowledge enrichment for the detection of missing knowledge. Different strategies have been defined to support the process, based on the similarity between entities stored into the INFINITY Knowledge base, discovering as result new "similarity patterns" that produce relationships among similar entities not yet connected. To this end, a set of similarity rules, built on entities handled in INFINITY and their metadata, has been defined in order to improve the accuracy of suggestions.

A case study to describe the approach focuses on entities of the type Person: every time a new entity of the type Person is stored into the INFINITY KG, similarity rules are applied by the Knowledge enrichment process to discover hidden relationships with entities of the same type already stored in the graph (as shown in Figure 7).

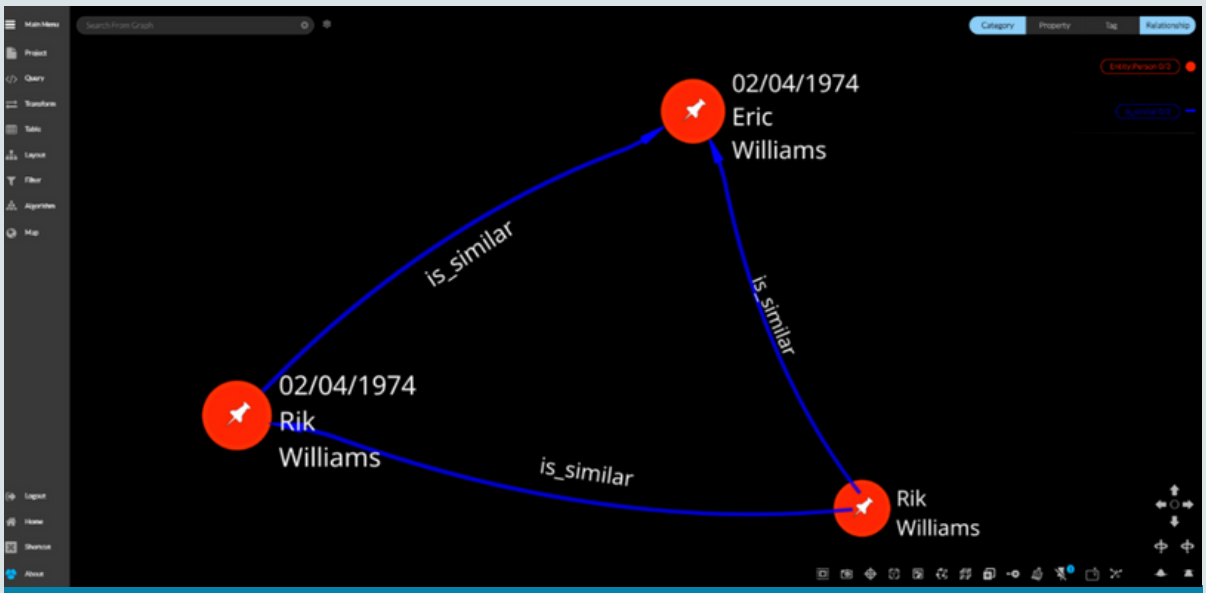


Figure 7: Example of generated similarity pattern

5. Future Opportunities

Ontologies provide a principled way to encode knowledge, by describing domain entities, their attributes and relationships. Through ontologies, the a priori knowledge can be formally detached from the application code, thus facilitating the design, development, and update of intelligent systems.

Ontologies offer a parsimony of concepts and are easily extendable, can be explanatory and, more importantly, can present hidden inferences because of the formal underpinnings. Specific reasoning mechanisms enable the possibility of:

- Detecting a new event or a suspicious activity by analyzing incoming resources and extracting information and merging it with previous knowledge;
- Inferring knowledge that is not immediately visible to the users, due to the large amount of information that could be acquired;
- Predicting new events on the basis of behavioral patterns detected in previous events;
- Providing suggestions and support to decision-making processes for LEA activities;
- Guiding users to search specific types of information that are missing and that can improve future reasoning processes, boosting in this way the investigations.

References

1. This Technical Brief was prepared by ENGINEERING, Rome, Italy, as part of T10.5.
2. Cf. Najgebauer, Andrzej, et al. (2008), "The prediction of terrorist threat on the basis of semantic association acquisition and complex network evolution", in: Journal of Telecommunications and Information Technology, pp. 14-20.
3. Cf. Turner, Matthew D., Weinberg, David M. and Turner, Jessica A. (2011), A Simple Ontology for the Analysis of Terrorist Attacks
4. Cf. The Basic Formal Ontology Org. (BFO): http://basic-formal-ontology.org/; and Basic Formal Ontology Org. (2020), BFO – Specification and User's Guide: http://www.ifomis.org/bfo/documents/manual.pdf.
5. Cf. Inyaem, Uraiwan, et al. (2010) "Construction of fuzzy ontology-based terrorism event extraction", in: Knowledge Discovery and Data Mining, WKDD'10. Third International Conference on IEEE, pp. 391-394
6. Cf. Donalds, Charlette M. and Osei-Bryson, Kweku-Muata A. (2019) "Toward a cybercrime classification ontology: A knowledge-based approach", in: Computers in Human Behavior, Vol. 92, pp. 403-418.
7. Cf. Barn, Ravinder and Barn, Balbir (2016) "An Ontological Representation of a Taxonomy for Cybercrime".his Technical Brief was prepared by DFKI, German Research Center for Artificial Intelligence GmbH, Department Augmented Vision, Kaiserslautern, as part of T10.5.