Invited keynote on IOT4SAFE 2020: Semantic Web Technologies in Fighting Crime and Terrorism: The CONNEXIONs Approach

Alexandros Koufakis, Despoina Chatzakou, Georgios Meditskos, Theodora Tsikrika, Stefanos Vrochidis, and Ioannis Kompatsiaris

Information Technologies Institute, Centre for Research and Technology Hellas
{akoufakis, dchatzakou, gmeditsk, theodora.tsikrika,
stefanos,ikom}@iti.gr

Abstract. Ontologies play a key role in the Semantic Web, providing the machine-interpretable semantic vocabulary and serving as the knowledge representation and exchange vehicle. On top of ontologies, various functionalities can be supported, such as semantic integration, enrichment and reasoning to either further enhance or enrich them with additional information, or to deduct implicit knowledge out of the already annotated information. This paper presents a holistic semantic model employed within the CONNEXIONs EU-funded project that is aimed at semantically representing and reasoning about all pertinent notions derived from the analysis of high volumes of heterogeneous data with the goal to ultimately improve the capabilities of Law Enforcement Agencies in their fight against crime and terrorism. The proposed model enables the compound of various important aspects that resolve around several sources of information considered important in this context, including online sources that are often adversely exploited, as well information produced by Internet of Thing devices, such as sensors and cameras.

Keywords: Semantic structures · Ontologies · Semantic Integration · Semantic Enrichment · Semantic Reasoning.

1 Introduction

The continuous evolution of serious and organized crime and terrorism poses significant challenges to Law Enforcement Agencies (LEAs) that need to be equipped with effective and efficient tools and solutions for the detection, prediction, investigation, and ultimately prevention of such activities. One particular challenge highlighted by the recent terrorist attacks in many European cities[1] is the growing role that online channels (e.g., Surface, Deep, and Dark Web and social media platforms) play towards enabling and facilitating terrorist causes, including radicalization, recruitment, and training, as well as the planning and coordination of terrorist attacks [14]. Thus, there is a need for LEAs to monitor

[1] https://goo.gl/daFnYZ

online sources so as to be in a position to detect, as early as possible, criminal and/or terrorist activities, in order to prevent and mitigate potential threats.

Moreover, during specific missions and operations undertaken by LEAs, it is of high importance for police officers (particularly those on the field) to be constantly aware of the overall situation, and in particular of any potential threats. For instance, when policing large-scale events, such as music festivals, it would be extremely important for LEA personnel to be aware of any suspicious activities taking place, since such information would allow them to act faster and possibly prevent a potential terrorist action. For instance, the monitoring and analysis of data produced from Internet of Things (IoT) devices, such as wearable and fixed sensors and cameras, could also assist towards this direction.

Therefore, the abundance of information produced from both IoT devices and online sources necessitates the development of methods and tools for efficiently representing processed and analyzed data based on a uniform modeling process. To this end, semantic structures (i.e., *ontologies*) can be utilized, which permit the semantic representation of all pertinent knowledge derived from the already analyzed data. Overall, ontologies permit the understanding, sharing, and reuse of knowledge across different systems, while also supporting various semantic operations, such as semantic integration, enrichment and reasoning, to either further enhance the available information or to infer logical conclusions out of the already semantically annotated information.

In this paper, we discuss how Semantic Web technologies can be used to effectively capture, interpret, and reason about information that revolves around criminal and terrorist activities, including evidence obtained both from online environments and also from sensors. The proposed approach has been developed within the context of the CONNEXIONs EU-funded project (`https://www.connexions-project.eu/`) and consists of an ontology and a range of semantic-based functionalities. The overall objective is to provide intelligence to LEAs that would allow them to improve their ability to analyze evidence and investigate crime and terrorism with the maximum possible effectiveness and efficiency.

2 CONNEXIONs Overview

CONNEXIONs aims to develop a next-generation platform that will assist LEAs to improve their capabilities to gather intelligence, analyze evidence, and thus investigate crime and terrorism in an effective and efficient manner.

The first step towards investigating crime and terrorism effectively and efficiently is to collect data related to a specific case. CONNEXIONs considers data obtained from various online sources, including the Surface, Deep, and Dark Web, and social media platforms, as well as data obtained from other sources, such as IoT devices and police reports; in addition, it can also consider digital evidence obtained from seized devices. This allows LEAs to acquire a more comprehensive view of potentially criminal and terrorism activities, thus enabling them to increase their chances to ultimately prevent them in time.

The next step constitutes the integration and correlation of such information through advanced methods (e.g., through ontologies and their supported functionalities) to finally deliver pertinent information in an interactive manner to a variety of end users, such as field officers, investigators, and analysts. To this end, immersive environments (such as virtual and augmented reality) are also considered for the delivery of such information in order to improve situational awareness, investigation, and training capabilities.

It is evident from the above that the establishment of a well-defined data representation framework is of high importance, since it permits the capturing of several modalities and thus enables the subsequent application of a variety of analytics tools. The following sections describe the knowledge structures of the CONNEXIONs ontology and the reasoning services that allow the deduction of logical inferences out of the already semantically annotated information (i.e., facts and relationships).

3 The CONNEXIONs Ontology

Ontologies are the starting point and the key component of Semantic Web technologies as they model the pertinent knowledge of any studied system. One definition states "*an ontology is a formal, explicit specification of a shared conceptualization that is characterized by high semantic expressiveness required for increased complexity*" [4]. In short, they provide the mold for the available data in order to build a homogeneous data repository that can effectively utilize Semantic Web technologies. In fact, such technologies are powerful tools towards enhancing decision support and reasoning capabilities towards fighting crime and terrorism. This section first discusses related work (Section 3.1), then outlines the approach adopted by the CONNEXIONs Ontology (Section 3.2), and finally presents the conceptualization of the proposed ontology (Section 3.3).

3.1 Related Work

Turner [12] proposed the Adversary-Intent-Target (AIT) ontology, namely a model for semantically representing adversary groups and their intentions, a classification of their weapons and attack types, and the relationship between the outcomes of an attack and the various recognized intentions of the adversary group. Another prominent approach is the work by Mannes and Golbeck [8,9], which proposes an ontology for representing terrorist activity and addresses the key issues the authors encountered during the development of the ontology, mostly revolving around how sequences of events can be described, and also representing the social networks that underpin terrorist organizations.

Moreover, Benahmed et al. [1] proposed an ontology for automating the characterization and the classification of terrorist threats at early stages, aiming at a more efficient threat mitigation, while Galjano et al. [5] focused more on monitoring subjects and objects (targets) of potential interest in an effort to monitor terrorist threats. In addition, an ontology for uncovering terrorism-related hidden semantic associations, which was the result of a knowledge fusion from several existing ontologies and open knowledge systems on the Web, was proposed by Chmielewski et al. [3].

Other relevant approaches include the one by Inyaem et al. [7], where a fuzzy ontology (whereby relationships have degrees of membership) for representing terrorism events was proposed. Also, Najgebauer et al. [10] developed a terrorism ontology for representing terrorist threat indications and facilitating the early detection of terrorist action preparation activities, constituting the backbone of an early warning system. Finally, Veerasamy et al. [13] presented an ontology specifically developed for cyberterrorism, which is aimed at identifying whether a cyber-event can be classified as a cyberterrorist attack or a support activity, and provides a rich semantic representation of underlying relationships, interactions, and influencing factors.

The main limitation of the ontologies presented above lies in their narrow focus, which typically revolves around criminal/terrorist groups and their intentions, attacks, and impacts. Though absolutely essential, these aspects constitute only a fragment of the knowledge required during the detection, prediction, and investigation of criminal and terrorist activities. The CONNEXIONs ontology, on the other hand, aims to provide a more holistic semantic model, covering various additional aspects besides the aforementioned ones, such as online behaviors, knowledge derived from the analysis of sensors-based data, as well as spatiotemporal information.

3.2 The CONNEXIONs Approach

The CONNEXIONs ontology aims to provide the means to model the relevant knowledge regarding the studied system, in order to enhance the decision support and reasoning capabilities for fighting crime and terrorism.

One important feature of the ontologies is their interoperability which is ensured by using common ontologies (and vocabularies) or by semantically mapping entities between different ontologies. Interoperability enables sharing knowledge from different repositories that are expressed using the same semantics. In practice, it is beneficial to reuse ontologies and vocabularies that are established provided they meet the particular needs.

Specifically, the CONNEXIONs ontology reuses the following resources about spatial information and multimedia modeling:

1. The Basic Geo (WGS84 lat/long) Vocabulary [2] is a lightweight ontology (formalized in OWL 2) that follows the specification of world geodesic system (revision 1984) for the representation of coordinates and altitude. This vocabulary directly supports only points in space that are defined by their coordinates. However, it can be extended to manipulate aggregations of points and subsequently form complex shapes.
2. The SIMMO (Socially Interconnected MultiMedia-enriched Objects) [11] is a model for expressing multimodal data in a social context with rich context. In particular, it focuses on their structure, interconnections and provenance

which renders it highly useful and versatile in expressing online user activity and media. In order for the model to be imported in the CONNEXIONs ontology it was first ported as an OWL 2 ontology.

The CONNEXIONs ontology is formalized in OWL [6], which is recommended by W3C and is designed to model complex knowledge systems. OWL has rich expressivity and can capture intricate relationships between entities. Moreover, OWL is logic-based language and supports implicit knowledge extraction and consistency verification.

3.3 The CONNEXIONs Ontology Conceptualization

The specification of the ontology is the first step towards exploiting semantic web technologies towards a reasoning framework that will aid the LEAs on crime event situations. In the following section the most important classes and properties of the CONNEXIONs ontology are presented (see also Figure 1).

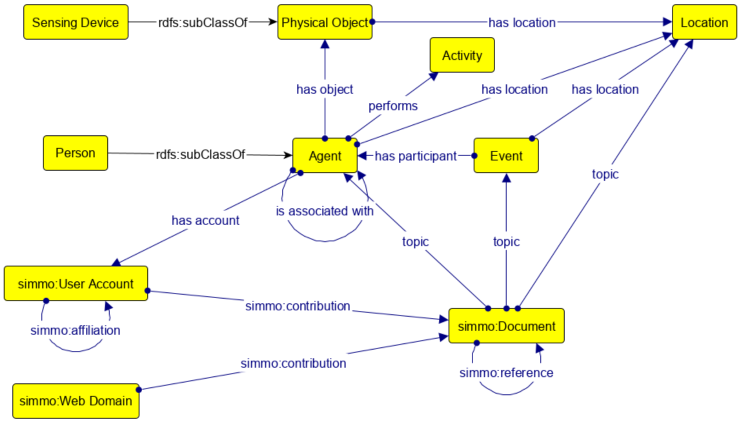

Fig. 1. The core classes of the ontology

The class *Event* represents a situation that takes place and is of relevance to the system; e.g., a social event, a crime or even an online event. This situation typically involves multiple entities that have varying roles and affect differently the event. Such entities with active role within the frame of the event are called *Agents* and their main characteristic is that they have the ability to affect the Event. Agents may be living beings (e.g., *Persons*) or web agents, namely, software entities that act on the Web (e.g., *User Account*). Additionally, Agents can perform activities that carry some significance for the event (the *Activity* class). Finally, agents might possess *Physical Objects*, such as sensors (*Sensing Device*). Physical objects can be carried or be stationary, where in both cases, the specification of their location is important (*has location* property).

Beyond the agents and the physical objects present at an event, online activity can provide a more complete view over the event. To this end, the ontology adopts the extended version of SIMMO model to express such online activities. In particular, the classes *User Account* and *Web Domain* are used to identify relevant online activities. In order to represent the activities and the artefacts posted online, the class *simmo:Document* is used in order to model multimodal and complex entities. Such documents can range from posts that contain a single text passage to web pages that contain multiple images and videos. Moreover, documents can reference (*simmo:reference*) other documents, or can be associated with other entities by a shared topic. User accounts are one of the more important contributors to the creation of documents and they can have relationships between them, e.g., a user account can be affiliated with others. Relationships between user accounts and documents are mainly expressed via the property *simmo:contribution* to express the creation of the document by the user.

Spatial information is crucial towards a complete understanding of a situation, especially in case of real world applications. For purely spatial information, the class *Location* is used that represents spatial information by reusing the Basic Geo Vocabulary that represents the locations as points in space. However, in time sensitive situations the spatial configurations are often fast changing, thus, a static spatial information is not sufficient. For this reason, we proposed the combination of temporal and spatial information using the *Spatio-Temporal Context* class (Figure 2), which is responsible for keeping a history of locations and their correspondent times.

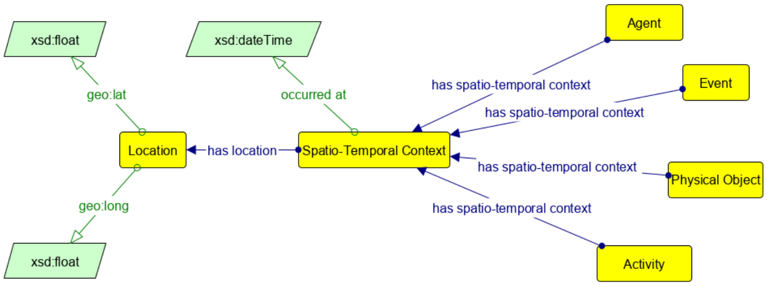

Fig. 2. Spatio-Temporal Context

4 Semantic Operations

Ontologies define a set of entities and the relationships between them at a theoretical level. However, in order to have practical applications these need to be emanated as a data structure that adheres to the semantic model defined by the ontology. In the context of the CONNEXIONs project, the data structure that holds the semantic information is an RDF database and comprises a Knowledge Base (KB) with an interface that utilizes the SPARQL (SPARQL Protocol and RDF Query Language) query language in order to insert and retrieve data. SPARQL is a semantic query language that operates on RDF data (W3C standard) and one of key technologies of the Semantic Web. At a higher logical level, the operations implemented with SPARQL are semantic integration, enrichment and reasoning. The first two are responsible for inserting knowledge into the KB, while the third for inferring hidden knowledge from the available knowledge.

4.1 Semantic Integration

Semantic integration is the process of consuming structured data that can be translated to semantic information and consequently be populated to the KB. Typically, the information is provided in the form of JSON (JavaScript Object Notation) data and the relevant information is extracted from it. In order for the information to be inserted to the KB it is transformed to SPARQL queries that update the contents of the KB. E.g., if face recognition was to performed on video footage obtained from some event, then the resulting information could include the following:

– Event ID, i.e., the ID of the social event where the person was recognized, e.g., “1234”.
– Person ID, i.e., the ID of the recognized person, e.g., “person 5262”.
– Timestamp, i.e., the moment when the person was detected, e.g., “2019-05-17T14:00:00Z”.
– The coordinates of the person, e.g., “latitude: 40.62” and “longitude: 22.94”.

In this case the corresponding SPARQL query that could be used to populate the new information to the KB would be similar to the presented in Listing 1.1.

```
INSERT {
    :person_5262 rdf:type :Person.
    :person_5262 rdf:type owl:NamedIndividual.
    :person_5262 :connID "5262".
    :loc_5262 rdf:type :Location.
    :loc_5262 geo:lat "40.62"^^xsd:double.
    :loc_5262 geo:long "22.94"^^xsd:double.
    :stc_5262 rdf:type :Spatio-TemporalContext.
    :stc_5262 rdf:type owl:NamedIndividual.
    :stc_5262 :hasLocation :loc_5262.
    :stc_5262 :occurredAt "2019-05-17T14:00:00Z"^^xsd:dateTime.
    :person_5262 :hasSpatio-TemporalContext :stc_5262.
    ?event :hasParticipant :person_5262.
    ?event_loc :containsLocation :loc_5262.
}
WHERE
{
    ?event :connID "1234".
    ?event :hasSpatio-TemporalContext ?event_stc.
    ?event_stc :hasLocation ?event_loc.
}
```

Listing 1.1. SPARQL query for semantic integration

4.2 Semantic Enrichment

Semantic enrichment is the process of enhancing the context of individual entities within the KB using a variety of sources. This process resembles semantic integration in that they both aim to populate the ontology. However, semantic enrichment typically refers to external sources, such as DBpedia (`https://wiki.dbpedia.org/`) and other public Knowledge Bases (e.g., criminal records).

In particular, let us examine the case of reporting a past violent activity against a VIP. This particular enrichment example ties with the integration discussed above and enriches the already known information. Data from the criminal records would include:

- Person (determined by their ID) who committed the violent activity, e.g., "5262".
- Target of the activity (ID), i.e., the entity that was intended to be harmed via the activity, e.g., "7382".
- Timestamp, the moment when the person was detected, e.g., "2017-04-12T14:00:00Z".
- The coordinates of the person, e.g., "latitude: 40.65" and "longitude: 23.01".

The SPARQL code that is presented in Listing 1.2 is produced given the previous information. The code is executed and the resulted changes are made to the KB. In this case, the entities that correspond to the event and the VIP participant were added in advance via integration methods, similar to the previous example.

```
INSERT {
    :act_0291 rdf:type :ViolentActivity.
    :act_0291 rdf:type owl:NamedIndividual.
    :loc_0291 rdf:type :Location.
    :loc_0291 rdf:type owl:NamedIndividual.
    :loc_0291 geo:lat "40.65"^^xsd:double.
    :loc_0291 geo:long "23.01"^^xsd:double.
    :stc_0291 rdf:type :Spatio-TemporalContext.
    :stc_0291 rdf:type owl:NamedIndividual.
    :stc_0291 :hasLocation :loc_0291.
    :stc_0291 :occurredAt "2019-04-12T14:00:00Z"^^xsd:dateTime.
    :act_0291 :hasSpatio-TemporalContext :stc_0291.
    :act_0291 :against ?vip.
    ?person :performs :act_0291.
}
WHERE
{
    ?person :connID "5262".
    ?vip :connID "7382".
}
```

Listing 1.2. SPARQL query for semantic enrichment

4.3 Semantic Reasoning

Semantic reasoning is the automated process of inferring implicit information from the knowledge available in the KB. Semantic reasoning is performed as a set of SPARQL rules that evaluate the necessary conditions for the inference to be valid. The inferences that result from the semantic reasoning are stored in the KB and can be forwarded to other tools if needed.

Semantic reasoning is performed automatically whenever the KB is updated via semantic integration and enrichment operations. In those cases, the semantic content changes and reasoning is applied to include the new information. Additionally, the semantic reasoning could be itself an enrichment process, by enriching the KB with new inferred information. In practice, semantic reasoning is triggered mainly by new analysis results becoming available. Moreover, a second method that triggers semantic reasoning is the use of specific reasoning requests from other tools or the end users. If this method is deployed as a second way to initiate reasoning, such reasoning request must be well defined. It is possible to allocate some reasoning methods to each of the two triggering mechanisms.

An example of semantic reasoning could examine whether a person that has performed violent actions against a VIP in the past attends the same social event as the VIP, then the person can be characterized as a potential threat. The SPARQL query (Listing 1.3) implements such a reasoning rule, and it is validated when the rule conditions are reached. Additionally, if the rule is realized, the appropriate updates are made to the KB.

```
SELECT ?person ?vip
WHERE {
    ?event :hasParticipant ?person.
    ?event :hasVIPParticipant ?vip.
    ?person :performs ?activity.
    ?activity rdf:type :ViolentActivity.
    ?activity :against ?vip.
}
```

Listing 1.3. SPARQL query for semantic reasoning

5 Conclusions & Future Work

This paper presented the Semantic Web technologies that are used in the context of the CONNEXIONs project towards an enriched reasoning framework for the support of LEAs in their fight against crime and terrorism. First, the ontology that acts as the basis for the semantic representation of multimodal heterogeneous and complex data, that are pertinent to the studied system, was presented. Next, the available semantic operations that aim to manipulate effectively the available data and provide useful insights about implicit knowledge that derives from the available data were illustrated. Overall, the presented framework is currently being successfully applied within the CONNEXIONs project and we aim to incorporate more complex semantic reasoning rules in the future.

Acknowledgments This work has received funding from the European Union's H2020 research and innovation programme as part of the CONNEXIONs (GA No 786731) project.

References

1. Benahmed, K., Assouli, N.: An ontology design for terrorism activities (2017)
2. Brickley, D.: Basic geo (wgs84 lat/long) vocabulary. Documento informal escrito en colaboración (2006)
3. Chmielewski, M., Gałka, A., Jarema, P., Krasowski, K., Kosiński, A.: Semantic knowledge representation in terrorist threat analysis for crisis management systems. In: International Conference on Computational Collective Intelligence. pp. 460–471. Springer (2009)
4. Feilmayr, C., Wöß, W.: An analysis of ontologies and their success factors for application to business. Data & Knowledge Engineering **101**, 1–23 (2016)
5. Galjano, P., Popovich, V.: Theoretical investigation of terrorism. ontology development. In: Information Fusion and Geographic Information Systems, pp. 227–239. Springer (2009)
6. Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P.F., Rudolph, S., et al.: Owl 2 web ontology language primer. W3C recommendation **27**(1), 123 (2009)
7. Inyaem, U., Meesad, P., Haruechaiyasak, C., Tran, D.: Construction of fuzzy ontology-based terrorism event extraction. In: 2010 Third International Conference on Knowledge Discovery and Data Mining. pp. 391–394. IEEE (2010)
8. Mannes, A., Golbeck, J.: Building a terrorism ontology. In: ISWC Workshop on Ontology Patterns for the Semantic Web. vol. 36 (2005)
9. Mannes, A., Golbeck, J.: Ontology building: A terrorism specialist's perspective. In: 2007 IEEE Aerospace Conference. pp. 1–5. IEEE (2007)
10. Najgebauer, A., Antkiewicz, R., Chmielewski, M., Kasprzak, R.: The prediction of terrorist threat on the basis of semantic association acquisition and complex network evolution. Journal of Telecommunications and Information Technology pp. 14–20 (2008)
11. Tsikrika, T., Andreadou, K., Moumtzidou, A., Schinas, E., Papadopoulos, S., Vrochidis, S., Kompatsiaris, I.: A unified model for socially interconnected multimedia-enriched objects. In: International Conference on Multimedia Modeling. pp. 372–384. Springer (2015)
12. Turner, M.D.: A simple ontology for the analysis of terrorist attacks (2011)
13. Veerasamy, N., Grobler, M., Von Solms, B.: Building an ontology for cyberterrorism (2012)
14. Weimann, G.: Terror on the Internet: The new arena, the new challenges. US Institute of Peace Press (2006)