An Ontology for Describing Security Events

Hossein Fani
University of New Brunswick
Fredericton, NB, Canada
hosseinfani@gmail.com

Ebrahim Bagheri
Ryerson University
Toronto, ON, Canada
bagheri@ryerson.ca

Abstract— Mining security events helps with better precautionary planning for community safety. However, incident records are expressed in diverse and application dependent formats which impedes common comprehension for automatic knowledge extraction and reasoning. In this paper, we present Security Incident Ontology, SIO, a novel light-weight domain ontology for security incidents. We use Timeline to annotate the temporal facts of incidents and adopt Event to represent any security issues from indecent behavior to assault to more adverse crime which raise the security alarm in a community. It will present a unique way to the security incident detectors, a police officer, Robocops, or intelligent CCTV cameras, to report security events. We use SIO in populating security incident notifications of Integrated Risk Management (IRM) at Ryerson University to evaluate its competency, for Ryerson University campus has both business and housing area in the vicinity and encompass not only high rate, but also wide variety of different security issues. SIO is developed in OWL 2 with Protégé.

Ontology; Semantic Web; OWL; Security Incident; Event.

I. INTRODUCTION

Environmental health and safety is the utmost priority of any municipal governor. In order to enhance the continual security of a community nearly real-time software systems are developed to monitor the area, record the events and send alarm notifications [1] [2]. For an ounce of prevention is worth a pound of cure, lots of efforts have been put to find security incidents’ root cause, prediction, and prevention by mining huge datasets of records [3]. Unfortunately, these records do not comply with a widely accepted standard in representation which impede the automatic knowledge extraction and reasoning. Local and official security guards or crime detector CCTV cameras [4] express same event with different schemes. In such a situation, we desperately need an ontology which, to our knowledge, thus far, we do not have any. In this paper, we devised a novel light-weight domain ontology, Security Incident Ontology (SIO), which is able to describe any kind of security risk from indecent behavior to assault to more adverse crime. Temporal aspects of security incidents are indispensable and SIO links to Timeline [5] ontology. Timeline is an extension to OWL-Time [6] [7] [8]. Timeline together with Event [9] ontology are able to address any general events. However, in SIO we specify the Event to security incident types and Agent class to Victim and Subject. SIO is developed in OWL2 with Protégé ontology editor. To show SIO capability in answering any competency question in security paradigm, we automatically extracted security incidents from security notifications of Integrated Risk Management (IRM) system at Ryerson University during year 2014 and represent them in SIO. Ryerson University campus has both business and housing area in its neighborhood and located at Toronto downtown. As a result, it can be a well suited area for not only high security threat rate, but also embodies wide variety of different crimes. According to our populated dataset, on a monthly basis, two security incidents of different types occur in the campus. Moreover, we are going to publish our populated dataset in Linked Data [10] to not only get it into the Linked Open Data Cloud [11], but also provoke SIO as the widely accepted representation for security events.

The remainder of this paper is organized as follows. Section 2 introduces the background and the initiatives for the work. In section 3, we construct event oriented security incident ontology. Section 4 presents the ontology evaluation by automatically populating security incident instances from Ryerson IRM notifications. Finally, the conclusions are given in section 5.

II. BACKGROUND

Ryerson University believes an informed community is a safer one. The Integrated Risk Management (IRM) system notifies all Ryerson staff, students, faculty and alumni (who have graduated within the past five years) by security incident alarms which are delivered directly via email [12]. For the urban campus is located at the downtown center of Toronto, the most populous, yet commercial capital city in Canada [13], such system seems indispensable to continually enhance the safety and security of the community. Each notification includes temporal facts of the incident, location, victim and suspect details, and a brief account of whole event. Likewise, Toronto Police Service (TPS) provides several mailing lists for which citizens of different divisions can sign up to be kept up-to-date on current happenings across the city, and in their community [14]. However, lack of standard way between reporting parties in representing security happenings and verification mechanism impedes automatic, yet reliable crime analysis and knowledge discovery for long-term planning. As a result, TPS has a disclaimer in its crime statistics webpage [15] and states firmly they make no warranty to the content, accuracy, timeliness or completeness of the statistics.

In order to provide a reliable, yet explicit understanding of incidents reported from vast variety of detectors we took an ontological approach. The ontology-based description approach is not novel. There are several initiatives to model event-focused concepts in knowledge representation [16], medical informatics [17] [18], sports [19], business news [20], scholarly events [21], life events [22] and multimedia community [23] [24]. Crime Emergency Event Model (CE2M) [25], despite what its name implies, is an ontology for emergency events

This work is funded by the Natural Sciences and Engineering Research Council of Canada.

DOI reference number: 10.18293/SEKE2015-101
rather than crimes. [26] constructs an event ontology to describe cyber-crimes on the level of event for crimes in Web such as online fraud, internet pornography, illegal trade, false advertising, violations of privacy, etc. And it is still not a general crime domain ontology. [27] describes a knowledge base of Politically Motivated Violent Events (PMVE) consisting of a domain ontology and of instance data. Although the work explain almost nothing about its ontology and focus more on extracting violent crimes from online news reports, it provides us, along with the aforementioned ontology models, an insight into how to model our generic security event structure. We reuse a most cited event ontology, Event [9], instead of recreating a new one and specifically adopt it to Security Incident Ontology. SIO, a novel explicit specification of security incident conceptualization.

III. SECURITY INCIDENT ONTOLOGY

A. Ontology Engineering

Ontology development is not an easy task. It requires skills and is still an art rather than technology. People need a sophisticated methodology to help them develop an ontology [28]. We use UPON, a methodology for ontology building derived from the Unified Software Development Process [29]. UPON is use-case driven and iterative process methodology, well-suited for developing domain ontologies. The iteration counts for inception, elaboration, construction, and transition phase are 1, 2, 2, and 1 respectively in the first release version of our ontology. Our ontology editor is Protégé. The environment not only facilitates reusing other ontologies by importing ontologies as well as its ontology library [30], but also has visualization tools which help an engineer to have big clear picture of ontology compartments. Finally, our ontology language is OWL2.

B. Event Ontology Adoption

Security incident is an event. That simply means:

```xml
<owl:Class rdf:ID="SecurityIncident"> <rdfs:subClassOf rdf:resource="#Event" /> </owl:Class>
```

We either can create a new upper level core ontology for event or reuse the most cited, yet suitable one for our work. We prefer the latter one and adopt Event [9]. This ontology is based on the view expressed by James F. Allen and George Ferguson in [31] which states that events are primarily linguistic or cognitive in nature and the world contains no events. Events are just certain useful and relevant patterns of world changes. Nonetheless, this ontology has already been proven useful in a wide range of context, due to its simplicity and usability. The ontology has `event:Event` at the heart and reuses Timeline [5] ontology for temporal predicate `event:time` and Geo RDF vocabulary [32] for spatial predicate `event:place`. Timeline ontology itself has OWL-Time [6] at heart to express instantaneous or extended time object along with a temporal algebra. Additionally, it is able to describe timelines other than the universal one which may be used in a recorded track or on any media with a temporal extent. `event:factor` is everything used in an event, `event:product` is whatever produced by an event and `event:sub_event` provides a way to split a complex event into simpler ones. Fig. 1.

However, security incidents have some discriminatory characteristics. The `event:agents` in Event ontology is a victim or suspect with this respect. `event:product` in security event is presumed as a crime or safety threat. Furthermore, an incident in this area may have spatial trajectory which demands change in `event:place` property. Thus, Event ontology should be customized to meet these needs.

C. Reification

From the survey and analysis of various security incidents, glossary of terms is formed. Table 1 are lists of new concepts included in our SIO and Fig. 2 shows its inter-relations. SIO leverage `event:agent` predicate of `event:Event` class for modeling the `sio:Victim` and `sio:Subject` of `sio:SecurityIncident`. We assume that security incident should have at least one `event:agent` of type `sio:Subject` which is `owl:subClassOf foaf:Agent` indirectly. This entity is the most enriched and has predicates to describe and track the subject in an incident such as, `sio:height`, `sio:weight`, `sio:hairColor`, `sio:image`, `sio:preState`, `sio:postState`. `sio:preState` explain the subject how he/she start to make a security concern and `sio:postState` shows the his/her final destiny e.g. flee, arrest, killed. `sio:Victim`, likewise, has these predicates. It is worth mentioning that on the one hand `sio:Subject` or `sio:Victim` may be `sio:CommunityMember`, and on the other hand there is no is-a or kind-of relationship between them. SIO leave it to the time when we instantiate a security incident. By then, we can add `rdf:type` to manage the security incident types. Moreover, a taxonomy of incidents has been defined. Each type of incidents has its own class.

In addition to the mentioned concepts, we found that a same incident can be reported by different report party, sometimes with time lag between them. Hence, we add similarity relation between incident entities to identify duplicated of a same incident. We eschewed the obfuscation of relativity of simultaneity [33] for the first version SIO and identify two incidents which are simultaneous in location of space (`event:place`) and location of time identical. As different source of detection fortify the incident integrity, we do not remove the duplicates and instead make an association between them by built-in OWL property `owl:sameAs`.

Figure 1. The Event Model. This ontology deals with the notion of reified events. It defines one main Event concept. An event may have a location, a time, active agents, factors and products.
The SIO ontology specification and its Turtle version are available at [http://semionet.rnet.ryerson.ca/ontologies/sio.owl](http://semionet.rnet.ryerson.ca/ontologies/sio.owl)

Table I. A list of the major concepts and entities (second level) in the SIO ontology version 1.0

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Security Incident</td>
<td>The set of connected events which reflects an occurrence, unusual problem, incident, deviation from standard practice, or situation that requires follow-up action.</td>
</tr>
<tr>
<td>Sexual Assault</td>
<td>A form of sexual violence, is any involuntary sexual act in which a person is threatened, coerced, or forced to engage against their will, or any non-consensual sexual touching of a person.</td>
</tr>
<tr>
<td>Assault with Weapon</td>
<td>In committing an assault, the subject carries, uses or threatens to use a weapon or an imitation thereof.</td>
</tr>
<tr>
<td>Assault</td>
<td>The direct or indirect application of force to another person, or the attempt or threat to apply force to another person, without that person’s permission.</td>
</tr>
<tr>
<td>Indecent Exposure</td>
<td>Indecent exposure is the deliberate exposure in public or in view of the general public by a person of a portion or portions of his or her body, in circumstances where the exposure is contrary to local moral or other standards of appropriate behavior.</td>
</tr>
<tr>
<td>Voyeurism</td>
<td>Voyeurism is the sexual interest in or practice of spying on people engaged in intimate behaviors, such as undressing, sexual activity, or other actions usually considered to be of a private nature.</td>
</tr>
<tr>
<td>Robbery</td>
<td>The act of taking or another person’s property (including attempts).</td>
</tr>
<tr>
<td>Victim</td>
<td>A victim of a security incident is an identifiable person who has been harmed individually and directly by the suspect, rather than by society as a whole.</td>
</tr>
<tr>
<td>Subject</td>
<td>A subject is a known person who initiate the security incident usually by committing crime.</td>
</tr>
</tbody>
</table>

Figure 2. A brief snapshot concepts reuse from foaf, geo, wn, time, event, and our sio. Red arrows imply inheritance or owl:subClassOf and the blues are associations.

IV. EVALUATION

We believe that a successful domain ontology should I) be capable of expressing any instance data and answer any competency questions in the associated domain, II) become popular and be reused by knowledge engineers in similar domains. To show the first one, we populate the security incidents which happens at Ryerson University urban campus. Integrated Risk Management (IRM) system at Ryerson notifies its community members including staff, students, faculty, and alumni (who have graduated within the past five years) about security incidents in a near real-time manner. The notification is delivered to the email which has a hyper link to the incident specific webpage. Ryerson could have provided us with the whole incidents as this work is done partly under Ryerson affiliation. Then we could easily extract information and transform it to SIO way of representation by an Extract-Transform-Load (ETL) engine or a mapper.

However, other engineers who wants to reuse our ontology may not have same chance of data accessibility. Hence, we continue to obtain the incidents information independently by crawling the incidents webpages. This way, we accompany our SIO with its crawlers as a fully-fledged solution. Moreover, we have provided SPARQL and JDBC endpoint to our dataset in jdbc:virtuoso://semionet.rnet.ryerson.ca:11111/charset=UTF-8/log_enable=2 and [http://semionet.rnet.ryerson.ca:8890/sparql](http://semionet.rnet.ryerson.ca:8890/sparql) with Graph IRI [http://ls3.rnet.ryerson.ca/SecurityIncident/test](http://ls3.rnet.ryerson.ca/SecurityIncident/test). That means we have satisfy the 5 steps of Linked Open Data. The steps (stars) are [34]:

- ★Available on the web (whatever format) but with an open license, to be Open Data
- ★★Available as machine-readable structured data (e.g. excel instead of image scan of a table)
- ★★★Non-proprietary format (e.g. CSV instead of excel)
- ★★★★Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
- ★★★★★Link your data to other people’s data to provide context
To provoke SIO reuse, as a future work, we show that our work is endorsed by the Semantic Web community by registering our dataset to Linked Open Data [11] and adding it to the Linked Open Data Cloud.

A. Automatic Instance Data Population

We develop an infrastructure to transform textual notification of security incidents to machine readable representation in SIO. The software system, namely Machine Readable Security Incident Notification (MRSIN), has two main components, Security Incident Service and Security Incident Data Mart. Fig. 3. The former consists of I) Natural Language Layer (NLL) which extracts temporal and factual information from incident notification text to data objects, II) Semantic Web Layer (SWL) which serializes objects to Web Ontology Language individuals, and III) Triple Access Layer (TAL) which stores the individuals to the well-known triple store, Virtuoso Universal Server. All layers are in compliance with layering architecture and are working within a passive run on service. The later component incorporates a geographical distribution data mart of security incidents to support police and government with decisions, and criminologists with suggestions. This layer plays as the user interface to the system.

NLL is supposed to parse the textual content of the security incident report and extract entities and their property-values along with the co-references. Extracting temporal expression such as incident report date and event date, finding the victim and subject of the event and their associated property-values are the major tasks in this layer. There are toolkits for this kind of task among which the Apache OpenNLP2 and Stanford University Stanford Named Entity Recognizer (NER)3 are the most cited. However, these libraries working properly in general text corpus and will fail in our domain specific context. Therefore, we should train them (the model) to learn our textual paradigm. Since this task make a research shift to the current one, we leave it for future work and stay with the NER (7 class model trained: Time, Location, Organization, Person, Money, Percent, Date) along with some customizations.

Java JDK 1.8 and PHP 5.5 are the programming languages for the service and data mart components respectively, NetBeans4 8.0 is the Integrated Development Environment (IDE). We use Apache Jena5 in our SWL and Virtuoso Jena Provider6 for TAL.

B. Knowledge Extraction

The ontology allows users to semantically search and retrieve security incident information. Examples of semantic search scenarios may be: finding incidents with a specific type of security threat, retrieving sub-incidents of an incident, or searching incidents in which one particular suspect is involved. These queries can be expressed by SPARQL query language and be asked from our dataset SPARQL or JDBC endpoint. We show a SPARQL query example in Fig. 4 to search different types of security incidents in a date range. The sample result would be such in Fig. 5.

---

1 http://virtuoso.openlinksw.com/
2 http://opennlp.apache.org
3 http://nlp.stanford.edu/software/CRF-NER.shtml
4 https://netbeans.org/
5 https://jena.apache.org/
6 http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VirtJenaProvider
V. RELATED WORK

There are several similar initiatives to model event-focused concepts in knowledge representation [16], medical informatics [17] [18], sports [19], business news [20], scholarly events [21], life events [22] and multimedia community [23] [24]. In [17] the authors present Adverse Events Reporting Ontology (AERO) to report adverse event, any untoward medical occurrence in a patient or clinical investigation. [18] designs an event ontology for application in the machine understanding of infectious disease-related events reported in natural language text. Concepts and terminology in the field of sports events and their relationships are studied in [19]. Pattern-based approach is used in [20] to build newsEvents ontology to model business events, the affected entities and relations between them. [21] presents a very thorough process from creating ontology, automatic data instance population, and knowledge extraction interface in domain of scholarly events. We highly appreciate this work and mainly follow the same approach in our work.

Much closer to our work’s domain we can name Crime Emergency Event Model (CE2M) [25]; although, despite what its name implies, it is an ontology for emergency events rather than crimes. Also, [26] constructs an event ontology to describe cyber crimes reflected in Web such as online fraud, internet pornography, illegal trade, false advertising, violations of privacy, network gambling, damage to reputation; it is still not a general crime domain ontology. [27] describes a knowledge base of Politically Motivated Violent Events (PMVE) consisting of a domain ontology and of instance data. Sadly, this work does not publish its ontology specification.

Barring [20], all aforementioned works create their own event ontology for their domain of study, despite the fact that we have capable event ontologies. Event [9] and TimeML [35] are two cases based on different ontological views of the world. The latter employ linguistics and uses mark-ups to express event and time entities while the former prefer to focus on real happening of the event and comply with RDF/N3 representation. Event also reuse the Timeline [5] for temporal references in events. Timeline reuse, also, OWL-Time [6]. We respect reuse practice and not only adopt Event as the base ontology for security incidents instead of creating a new one, but also develop our SIO with regard to reuse principle.

VI. CONCLUSION

In this paper, we present Security Incident Ontology, SIO, a novel light-weight domain ontology for security incidents. We use Timeline ontology to express time references and adopt Event ontology to representing any security threats which raise the security alarm in a community. SIO is developed in OWL2 with Protégé by UPON ontology engineering methodology. SIO wants to present a unique ontological machine readable way to the security incident detectors to report security events. To reinforce the feasibility and capability of our work, we populating security incident notifications of Integrated Risk Management (IRM) at Ryerson University in SIO. To comply with the Semantic Web community principles we publish our dataset and provide SPARQL endpoint to extract knowledge. We hope that SIO wedge his way into the Web and obtain highest reuse rank to raise human and machine readability and common ubiquitous comprehension of security incidents.

ACKNOWLEDGMENT

The authors graciously acknowledge funding from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES


