Development of a Semantic Representation Model of Criminal Information to Support the Assessment of Risk Situations

Jordan F. Saran^{1,2}, Leonardo C. Botega^{1,2}

¹Programa de Pós-graduação em Ciência da Informação - PPGCI Universidade Estadual Paulista Júlio de Mesquita Filho - UNESP – Marília, SP – Brasil

²Grupo de Interação Humano-Computador - GIHC Centro Universitário Eurípides de Marília - UNIVEM – Marília, SP – Brasil

{jordansaran,botega}@univem.edu.br

Abstract. Situational Awareness (SAW) refers to the level of consciousness that an individual or team holds over a situation. In the area of risk management and criminal data analysis, SAW failures can induce human operators to make mistakes in decision making and pose risks to life or property. In this context, risk assessment processes, which commonly involve data mining, fusion and other methods, present opportunities to generate better information and contribute to the improvement of the SAW of crime and risk analysts. However, the characterization of complex scenarios is subject to problems of representation and expressiveness of the information, which may influence its interpretation due to their quality and significance, generating uncertainties. The state-of-the-art in representation of information on risk situations and related areas presents approaches with limited use of information quality. In addition, the solutions are restricted to syntactic mechanisms for the determination of relations between information, negatively restricting the assertiveness of the results. Thus, this paper aims to develop a new approach to semantic representation of information of risk situations, more specifically creating domain ontologies, instantiated with crime data and information quality. In a case study, real information on crimes, represented by the new semantic model and consumed by computational inference processes, was be processed, aiming to characterize robbery and theft situations.

1. Introduction

In risk management domain, critical system operators need to act decisively with limited time, with overload of information that is often incomplete, contains a high level of psychological pressure among other factors that act In a negative way in the perception and understanding of the situation. Such problems compromise the outcome in responding to a critical situation because decision making can be inaccurate. In this context, situational awareness (SAW) helps human operators of critical systems to measure and improve their levels of perception, understanding and projection of the situation. SAW is an important cognitive process for decision makers in several critical areas and concerns the perception of the presence and disposition of entities of interest in an environment, the understanding of the meaning and importance of their individual actions and In space-time and the projection of their status in the near future [Endsley 1988].

In risk assessment systems, the acquisition and improvement of SAW levels of human operators becomes a complex task, since such systems collect, process and represent data from primarily human intelligence (HUMINT). Such reports can produce data with low levels of quality because its sources are heterogeneous, unpredictable, complex and dynamic. When incomplete data is propagated in decision-making support systems, they produce inaccurate information, which can negatively influence the SAW of the human operator, which depends on the correct specifications to understand the situation and the taking of Assertive decision.

In this context, it was identified that, together with the need to adopt a methodology for assessing the quality of data, the use of ontologies as semantic models can help to improve the representation of information useful to the system and the end user.

To highlight this issue, works such as [Mescherin et al. 2013] use the ontology in conjunction with crises metamodels to increase interoperability and decrease the lack of standardization of vocabularies used in communication between risk systems or the work of [Carvalho et al. 2009] that uses probabilistic ontology (PR-OWL) to represent critical areas characterized by the presence of uncertainty. Both works use ontology to solve problems related to their specific domains, however, do not include the data quality factor as a way of qualification of information in their ontologies, to enrich the representation and Inference of information.

The work [Souza et al. 2018] proposes the construction of an accurate knowledge about traffic conditions and the levels of insecurity in urban roads, calculating faster and safer routes, through the use of trafficking conditions, associated with Mapping and analysis of criminal events. Observing the work [Souza et al. 2018] it would be possible to generate inputs on the traffic conditions from the semantic model, thus enabling better conditions in the calculation of routes.

Thus, this work aims to contribute to the development of the SAW of the users of the system of assessing risk situations using ontologies to represent the information in a clearer and more objective way, demonstrating to the human operator all Possible relationships, entities and their attributes with a risky situation, where criminal data will be used. He will also be using a quality management methodology to quantify and qualify the data used by him in order to help the inferences processes on the risk situation.

2. Development of the Semantic Model

This project follows the development methods and definitions of the JDL pattern of fusion and data representation (Joint Directors Laboratories) [Blasch et al. 2013], and the guidelines of the Quantifymodel of fusion and information representation (Quality-awareHuman-driven information Fusion Model) [Botega et al. 2017] (Figure 1), developed by the human-Computer Interaction Group (HCIG).

This model evidences the need to complement syntactic processes of evaluation of situations with the representation and semantic processing, fueled by the evaluation of the quality of data and information. To develop this work, the methodology applied are based on the following activities: (1) Requirements Survey: Research on the main topics exposed in this work as risk management, situational awareness, management of risk with criminal data, and development and application of a questionnaire to experts in the



Figure 1. Process of Evaluation of Situations of Risk for the improvement of the Situational Awareness

field; (2) Task analysis aimed at the objective: application of the task analysis directed by the technical Meta(GDTA); (3) Development of an ontology of criminal domain: use of the 101 methodology and the vocabulary DQV (Data Quality Vocabulary) to develop the ontology; (4) Application of Quality Management: use of the IQESA (methodology forassessing the quality of information in the context of the consciousness of the emergency situation), to apply the metrics and dimensions of the quality of the data; (5) Validation of ontology: Validation of ontology developed using the SPARQL qeury.

2.1. Requirements Survey

After a study of the main themes of this study, a questionnaire was developed and applied to the military police of the state of São Paulo (PMESP). With the answers obtained, it was possible to identify which information is a priority for decision-making in a crime situation, and what are the services and data sources that should be consumed to provide them or deduce them. With this entry, it was possible to construct an attribute tree, this feature becomes an important one for the other phases of development because, it provides the entire hierarchy of attributes through the nodes that compose it, and for the qualification phase of the data that provides inputs to identify requirements that must be quantified and qualified.

From these results, it was possible to update the first version of a goal-driven task analysis Model (GDTA). This model represents beyond the objectives to be achieved, decisions to be taken, tasks to be fulfilled and information requirements to contemplate this demand [Endsley 2001]. The level of importance of each information revealed by the interview also enabled the creation of a scale of informational priorities.

2.2. Goal-Directed Task Analysis (GDTA)

The use of the GDTA methodology [Endsley 1988] aims to identify the main aspects related to the domain being applied. In critical systems, it is used to identify the main objectives to be successfully achieved by the operator, as well as to contribute to the acquisition of SAW and development of such operator. The GDTA is based on goals or goals, and not on simple tasks, according to it, this is because the objectives are the basis for decision-making in many environments with a high level of complexity, such as the military area [Endsley 2001]. Figure 2 shows part of the GDTA developed from the attribute tree.



Figure 2. Goal-Directed Task Analysis (GDTA)

2.3. Development of the criminal domain ontology

The methodology adopted for the elaboration and development of ontology will be the 101 methodology [Noy and McGuinness 2001]. Following a development cycle divided into seven stages, this methodology became the most used in the construction of ontologies due to its flexibility and simplicity of being applied in any domain and easy to understand. The steps for the development of ontology in the present study were:

- **Define scope**: The domain of application and study of this ontology are situations of risk, in which for this work are situations of theft and theft exclusively, where their function will be assisting in the generation of richer and more meaningful information for the Development of SAW specialists.
- **Consider the reuse**: No ontology was found related to the same domain of this work that helped in the development of it, especially in the use of Portuguese language and the organization of criminal problems. However, it was identified an external ontology that can be applied in the construction, was the DQV used to apply quality indexes within an ontology (W3C).

- Enumerate terms: The execution of this step was through the analysis of the responses obtained from the questionnaire applied to operators of critical systems and public and private security in conjunction with an analysis of criminal records on official bases, thus obtaining a Survey of the main terms found in which they went through the validation of an expert.
- **Define Classes and properties**: Based on the questionnaire applied, it was possible to identify the most relevant information to characterize a situation of laughter in the criminal domain. Based on the result of this questioning, the main classes of ontology were defined: person, place, object, situation and EC (Criminal event). For each of the mentioned classes were created more specific subclasses as well as the properties that characterize them.
- **Define constraints**: After defining classes and properties, it was necessary to establish domain constraints, such as: "Situation at least there will be a location", which receives cardinality 1 N within the ontology.
- **Create instances**: The ontology was fed with several instances that characterize the information pertaining to the criminal domain, which are presented by Figure 3 after the previous steps.



Figure 3. Main classes identified by the methodology 101

2.4. Quality Management Methodology Application

For quality management in conjunction with the ontological development process, functions and metrics of the IQESA methodology were used [Botega et al. 2016]. In addition, the DQV (Data Quality Vocabulary) vocabulary was used, with the objective of identifying the relations between the quality dimensions and the ontological instances. To qualify and quantify the data, the dimensions were used: currentness, completeness, consistency, relevance and certainty. Currentness: It is a temporal progress of an event, its contribution to the development of SAW systems operators of situations assessment is essential due to the need for updated data [Botega et al. 2016]. Where θ consists of an index composed of the current time and the metric is obtained by two attributes presented in the situation of occurrence, being represented by the Σ of these attributes: (1) Time that the crime occurred and (2) Time that the report was performed. This evaluation has two values as a result: one of them being a quantitative indicator for the existence of two attributes and how many minutes have passed since the report emerged. Equation 1 was adjusted to perform the current quantitative count [BOTEGA 2016].

$$\sum_{y=1}^{2} -\theta \tag{1}$$

• Completeness: To evaluate the completeness of the criminal domain, a tree of attributes was used, because it has all the priority attributes presented to meet a case report. Equation 2 demonstrates how to perform weight calculation.

$$C = S[(\frac{\Sigma\beta \cdot \gamma}{\Sigma\gamma} \cdot 0.9) + 0.1]$$
⁽²⁾

Equation 2 calculates the completeness dimension for the attributes present in the object to be analyzed, and S represents the presence of the object to be evaluated, when present S = 1, and when absent S = 0; β represents the attribute that describes the object, when present equals 1, and when missing is 0; The γ represents the weight, which, when considered a priority, has a value equal to 2 and when no priority is received the equal value 1. For each present weight, the sum of the multiplication and its result divided by the total of priority attributes should be made. The resulting value of this formula should range from 0 to 100% [BOTEGA 2016].

- Relevance: There is no equation to calculate relevance, it consists in identifying the relevant attributes presented in the object, assigning 0 to the irrelevant attributes and 1 to the relevant ones.
- Consistency: In the context of this work, the consistency assessment will be applied from the second occurrence regarding the same situation as the system receives, and the value assigned for consistent occurrences will be from 0 to 100%. Below are examples of inconsistent occurrences.
 - Report 1: "The victim stopped the vehicle at a traffic light, was surprised by two individuals who were on a motorcycle twister, black color. The croup hit with a gun in the glass and utting serious threat, subtracted the victim's vehicle. The authors escaped with an ignored destination. Victim has no conditions to describe the authors with details. Requested the lock next to Cepol. Nothing more. "
 - Report 2: "Two guys on a dark-colored motorcycle stole a car at the lighthouse at the intersection of Mooca Street and Taquari, pointed a revolver and took the woman's car. One of them was in a red coat."
 - Report 3: "An individual stole a gray bike on Mooca Street at the intersection of Fernando Falcão Street. The perp was armed and left sense to the Parque da Sabesp."

The three occurrences were obtained by PMESP, the first two are consistent because both refer to the characteristics of the criminals, one being reported to have seen two individuals on one motorcycle and the other that the motorcycle was black, the third occurrence is inconsistent due to the location where the crime occurred and the stolen goal being different from the other two occurrences.

• Certainty: At the end of the qualification and quantification of all dimensions and attributes, the assessment of the certainty of the situation is made. This dimension consists in determining a global index to represent the assessment of the quality of information as a whole.

2.5. Ontology Validation

To validate the present ontology developed throughout this work will be presented a case study in this section as an objective to demonstrate the quality of data and information, together with the ontology, as a support for critical systems and, in consequence, to help the operators of these systems to improve their SAW, promoting high level abstractions and reducing uncertainties at the time of decision making. To demonstrate such results, the IQESA methodology will be used along with the ontology the DQV vocabulary that represents the quality metrics and dimensions.

A situation of robbery will be addressed in this context, it will be necessary to understand the scenario of the crime situation and to identify some information relevant to the resolution of the case, such as where the crime was held, how many criminals had during the fact, What types of clothing the criminals were using during the crime and what kind of object was stolen. The following is the first occurrence used for this study:

Report 1: "The victim stopped the vehicle at a traffic light, was surprised by two individuals who were on a motorcycle twister, black color. The croup hit with a gun in the glass and uttering serious threat, subtracted the victim's vehicle. The authors escaped with an ignored destination. Victim has no conditions to describe the authors with details. Requested the lock next to Cepol. Nothing more. "

With this information obtained from the report it is possible to extract the information such as the time of the crime, the location, stolen object, characteristics of the victim among other information. After this first contact with the occurrence, the system itself in a data merger process organizes the information based on the process of assessing risk situations illustrated in Figure 1.

Entities	Completeness	Relevance	Currentness	Consistency	Certainty
Local	0%	0	-	-	-
Occurrence	90%	3	-	-	-
Victim	91,82%	4	-	-	-
Criminal	30%	1	-	-	-
Crime	100%	1	-	-	-
Object	100%	1	-	-	-
EC	100%	1	-	-	-
Situation	85%	4	375 min	-	28,42%

Table 1. QUALITY ASSESSMENT 1° REPORT

In a next step, the IQESA methodology is used to qualify and quantify the data and information along with the Completeness, Relevance, Currentness and Consistency dimensions presented in Table 1.

The Consistency dimension is calculated only from the second occurrence on the same crime situation. The calculation of certainty is given by the sum of all the calculated indexes (592%), divided by the amount of incides in the report(29) resulting in 28.42%. The Completeness, Relevance and Currentness indices are calculated according to the formulas shown in the previous sections:

- Occurrence: Of the 6 attributes present in the class, only 5 have values, applying these values the formula of completeness present in Equation 1, the index found is 90%. Of the 5 attributes present, only 3 are relevant.
- Crime: It has only one attribute as value, so the completeness index is 100%. It has a value of 1 relevance.
- **Object**: It has only one attribute as value, so the completeness index is 100%. It has a value of 1 relevance.
- EC(Criminal Event): It has only one attribute as value, so the completeness index is 100%. It has a value of 1 relevance.
- Victim: Of the 7 attributes present in the class, only 6 has value generating a completeness index of 91.82%. Of the 6 attributes present, only 4 are relevant.
- **Criminal**: Of the 6 attributes present in the class, only 1 has value generating a Completeness index of 30%. of the 1 attribute present, only itself is relevant.
- Situation: Of the 7 attributes present in the class, only 6 has value generating a Completeness index of 82%. Of the 6 attributes present, only 4 are relevant.

To calculate the currentness, the time the crime described by the victim, in this case at 11:30, was subtracted from the time the victim performed the report at 18:15, thus generating a current value of 375 minutes. The second report of occurrence was made by an anonymous complaint, being done after a few minutes of the crime.

Report 2: "Two guys on a dark-colored motorcycle stole a car at the lighthouse at the intersection of Mooca Street and Taquari, pointed a revolver and took the woman's car. One of them was in a red coat. "

It is possible to verify that this report contains more information about the situation in general, e.g. information about the location where the crime occurred and whether the criminals had some kind of firearm. Each new information, referring to the same situation as the system receives, occurs the same process of evaluating the quality of the data. Table 2 presents the new results regarding the dimensions of certainty, completeness, consistency, timeliness and relevance for the situation.

In Table 2 it is possible to obtain consistency, this measure is used to process the information for a situation that is already in progress, to apply a metric that is needed a new information base for a parameter that arrives at the system. he calculation resulted in 100% because the two occurrences contained information about the same crime, some even repeated confirming its consistency. The third report of occurrence was made by an anonymous complaint, being done after a few minutes of the crime equal the second report of occurrence.

Entities	Completeness	Relevance	Currentness	Consistency	Certainty
Local	26%	6	-	-	-
Occurrence	82%	3	-	-	-
Victim	82%	4	-	-	-
Criminal	46%	4	-	-	-
Crime	100%	1	-	-	-
Object	100%	1	-	-	-
EC	100%	1	-	-	-
Situation	91%	4	5 min	100%	36,13%

Table 2. QUALITY ASSESSMENT 2° REPORT

Report 3: "Two men robbed a gray car in the street from Mooca to the side of Santander bank . The men were on a black motorcycle one of them was in a blue jacket and jeans. They left in the direction of the villa lobos hospital."

This last occurrence presents some more important information, in which better characterizes the current situation of the crime as emphasize that the place where the crime could have occurred was in "Mooca street near Santander bank" and that the amount of Criminals who have performed such an act, in fact, are "two misers".

Entities	Completeness	Relevance	Currentness	Consistency	Certainty
Local	81%	6	-	-	-
Occurrence	82%	3	-	-	-
Victim	82%	4	-	-	-
Criminal	46%	4	-	-	-
Crime	100%	1	-	-	-
Object	100%	1	-	-	-
EC	100%	1	-	-	-
Situation	100%	4	8 min	100%	44,55%

Table 3. QUALITY ASSESSMENT 3° REPORT

In Table 3 It is possible to see again the calculation of consistency, where it obtained the result of 100% because the three occurrences contained close information that characterizes a single situation. The Table 4 illustrates the quality values measured for the fusion of occurrences 1, 2 and 3, the process for these calculations were the same as previously shown.

After these calculations on the quality, the ontology was instantiated and demonstrated in Figure 4. Since the ontology has already been conceptually considered a mechanism of representation of information, by being instantiated with the data present in the three occurrences the ontology may present new relationships between the entities, this type of process contributes positively in Inference of knowledge to the human operator of risk assessment systems.

To validate the ontology presented at Figure 4 was utilized SPARQL queries, these queries permit the user of the risk assessment systems a better inference about the crime situation, generating more certainty for an assertive decision-making. Consider the pre-

Entities	Completeness	Relevance	Currentness	Consistency	Certainty
Local	81%	6	-	-	-
Occurrence	82%	3	-	-	-
Victim	92%	4	-	-	-
Criminal	46%	4	-	-	-
Crime	100%	1	-	-	-
Object	100%	1	-	-	-
EC	100%	1	-	-	-
Situation_1	82%	4	375 min	100%	-
Situation_2	91%	4	5 min	100%	-
Situation_3	100%	4	8 min	100%	-
Situation	-	-	388 min	100%	70,28%

Table 4. QUALITY ASSESSMENT 1,2 AND 3° REPORT



Figure 4. Instantiated ontology considering reports 1, 2 and 3

fixes demonstrated at Source Code 1 for the queries presented in this section.

Source Code 1. Prefixes used in SPARQL queries

PREFIX rdf: <http://www.w3.org/1999/02/22 - rdf-syntax-ns#> **PREFIX** rdfs: <http://www.w3.org/2000/01/rdf-schema#>

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dqv: <http://www.w3.org/ns/dqv#>
PREFIX crimevoc: <http://localhost:7200/crimevoc#>
```

The first query performed shows information about the Local class, where they are referring to the address, latitude, and longitude that are extremely important in the criminal domain, because considering this information the resources are allocated. The Table 5 demonstrates the results obtained by the query shown at Source Code 2.

Source Code 2. First query

```
SELECT *
WHERE {
          ?situation crimevoc:happens ?crime .
          ?crime crimevoc:occurs_in ?local .
          ?local crimevoc:type_street ?type_street .
          ?local crimevoc:street ?street .
          ?local crimevoc:number_street ?number_street .
          ?local crimevoc:complement ?complement .
          ?local crimevoc:point_reference ?point_reference .
          ?local crimevoc:neighborhood ?neighborhood .
          ?local crimevoc:state ?state .
```

```
}
```

Atribute	Value
Situation	crimevoc:robbery
Crime	crimevoc:situation
Local	crimevoc:local
type_street	"Street"
street	"Mooca"
number_street	,,,,
complement	,,,,
point_reference	"Santander bank"
neighborhood	"Mooca neighborhood"
city	"São Paulo"
state	SP

Table 5. RESULT OF THE FIRST QUERY.

The second query was performed in the Criminal class, the information shown below refers to the types of objectives used by criminals during the crime with the values referring to the dimensions of Completeness and Relevance, provided by data qualification. The Table 6 demonstrates the results obtained by the query shown at Source Code 3.

Table 6. RESULT OF THE SECOND QUERY.

Atribute	Value	
Object	Car	
Completeness	92%	
Relevance	4	

The third query was performed in the Situation class, the information shown below refers to the percentage of certainty that all information contained in the ontology together with the data quality obtained during the construction of a single situation. The Table 7 demonstrates the results obtained by the query shown at Source Code 4.

Source Code 4. third query

```
SELECT *
WHERE {
    ?situation crimevoc:has ?graph .
    ?graph dqv:hasQualityMeasurement ?qt .
    ?qt dqv:isMeasurement ?Metric .
    FILTER regex(str(?Metric), 'Metric(SituatuionCertainty)'
    , 'i') .
    ?qt dqv:value ?value .
}
```

Table 7. RESULT OF THE THIRD QUERY.

Atribute	Value
Certainty	70,28%

3. Conclusion

The present work had the goal of demonstrating the development of a criminal domain ontology aware of data quality that will provide subsidies to achieve the different levels of situation awareness by the operator in risk management systems. However, the ontology developed incidentally, can only be used to solve crimes of robbery, theft and extortion by kidnapping.

The use of ontologies together with risk assessment systems aims to provide the system with an inference model in which computational processes can hold a greater knowledge about the situation and the relations, this happens when the data used have quality. To the human operator of risk management systems the benefit of the ontology with quality data next to the system extends both the construction of smarter interfaces with more accurate and better represented information as well as the transmission of greater knowledge of the situation even with the dynamicity that it possesses, thus promoting a higher level of trust to the human operator in decision making.

In order to illustrate the results obtained in this work, a case study was developed and presented, in which it can be verified that through queries that meet the operator's needs and the situation he is attending at any given moment, the ontology together with the quality management can contribute positively to the human operator, thus making it possible to identify explicit and non-explicit relationships between the entities present in the situation, such as identifying which object in the situation belongs to the criminal and the victim.. Such knowledge and control of the situation contributes positively to the acquisition and maintenance of SAW levels of such operators, thus allowing safer and more assertive decision making.

References

- Blasch, E., Steinberg, A., Das, S., Llinas, J., Chong, C., Kessler, O., Waltz, E., and White, F. (2013). Revisiting the JDL model for information exploitation. *Proceedings of the* 16th International Conference on Information Fusion, FUSION 2013, pages 129–136.
- BOTEGA, L. C. (2016). MODELO DE FUSAO DIRIGIDO POR HUMANOS E CIENTE DE QUALIDADE DE INFORMACAO. PhD thesis, Universidade Federal de São Carlos(UFSCar), São Carlos.
- Botega, L. C., de Souza, J. O., Jorge, F. R., Coneglian, C. S., de Campos, M. R., de Almeida Neris, V. P., and de Araújo, R. B. (2016). Methodology for Data and Information Quality Assessment in the Context of Emergency Situational Awareness. Universal Access in the Information Society, 16(4):889–902.
- Botega, L. C., Valdir, A. P., Oliveira, A. C., Saran, J. F., Villas, L. A., and De Araujo, R. B. (2017). Quality-aware human-driven information fusion model. 20th International Conference on Information Fusion, Fusion 2017 - Proceedings.
- Carvalho, R. N., Laskey, K. B., Costa, P. C. G., Ladeira, M., Santos, L. L., and Matsumoto, S. (2009). Probabilistic ontology and knowledge fusion for procurement fraud detection in Brazil. *CEUR Workshop Proceedings*, 527:3–14.
- Endsley, M. R. (1988). Design and Evalutation for Situation Awareness Enhancement. *Proceedings of the human factors and ergonomics society annual meeting*, 32(2):97–101.
- Endsley, M. R. (2001). Designing for Situation Awareness in Complex Systems. *Proceedings of the 2nd International Workshop on Symbiosis of Humans, Artifacts and Environment*, pages 1–14.

- Mescherin, S. A., Kirillov, I., and Klimenko, S. (2013). Ontology of emergency shared situation awareness and crisis interoperability. *Proceedings - 2013 International Conference on Cyberworlds, CW 2013*, pages 159–162.
- Noy, N. F. and McGuinness, D. L. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. *Stanford Knowledge Systems Laboratory*, page 25.
- Souza, A. M. D., Botega, L. C., Garcia, I. C., and Villas, L. A. (2018). Por Aqui é Mais Seguro : Melhorando a Mobilidade e a Segurança nas Vias Urbanas.