User-Driven Methodology for Data Quality Assessment in the Context of Robbery Events

Jessica Souza[1], Leonardo Botega[1,3], José Eduardo Santarém Segundo[2], and Claudia Berti[3]

[1] Centro Universitário Eurípides de Marília – UNIVEM, Marília, Brazil
{osz.jessica,leobotega}@gmail.com
[2] Universidade de São Paulo – USP, Ribeirão Preto, Brazil
santarem@usp.br
[3] Universidade Federal de São Carlos - UFSCar, São Carlos, Brazil
claudiabberti@gmail.com

Abstract. Situation assessment is a significant process for Situation Awareness (SAW) improvement since this process can be degraded if obtained by means of low quality data, which undermines the decision-making process. Identification of situations in military systems is of utmost importance since the information acquired is used in order to answer emergency calls in which poor understanding of situations for further resources allocation to answer it impacts on the victims involved. In order to assist the first level of SAW formation in the context of military situation assessment of robbery, the present work aims to present a methodology to provide support to the development of military decision-making systems addressed to represent and evaluate data quality. Such methodology is composed by gathering requirements for the robbery domain, the establishment of quantitative metrics for quality evaluation such as completeness and timeliness, the development of an ontology to capture and represent the semantic knowledge and the enforcement of evaluation functions according to scores set by means of quality metrics. Finally, it will be addressed a case study in response to robbery reports to illustrate the applicability of the methodology.

Keywords: Information Management, Data quality methodology, Situation Awareness, Knowledge management.

1 Introduction

Situation Awareness (SAW) consists in the human cognitive process of being aware of the situation that involves the user, and it is composed by three levels: perception of the elements in the environment, understanding the state of these elements, and the evolution of these in a near future. Achieving complete SAW is a process that takes place in the human mind and it is extremely important to capture knowledge when operating decision-support systems, as well as developing systems that can help this process.

© Springer International Publishing Switzerland 2015

Á. Rocha et al. (eds.), *New Contributions in Information Systems and Technologies*,
Advances in Intelligent Systems and Computing 353, DOI: 10.1007/978-3-319-16486-1_16

Mental model and SAW are highly related, since a mental model is formed by the user's understanding of situations. Hence, a poor understanding of situations affects the user's mental model, which leads to poor comprehension and jeopardizes decision-making. SAW Systems can provide several barriers to understanding, such as critical aspects and dynamics and real-time, in progress crime call answering of military police of São Paulo State in Brazil.

Saw is a process that requires cognitive activity and the prior user mental model of situations can assist to reduce the cognitive overload in this activity. However, poor understanding of information may not only cause the loss of its global significance, but also lead to failures while allocating resources.

Another important factor that can harm a user's mental model is the quality in which such complaints describe the reality of the scenario, since the information quality acquired in this process can lead to uncertainty to the operator when attempting to interpret the data.

The literature registers approaches that aim to enhance and maintain the user's SAW. This activity employs technologies such as cognitive models, ontologies and frameworks based on core ontologies, fuzzy logic, and data fusion models.

A way to provide support to both information quality and decision-support systems are ontologies, which can adapt to specific contexts and create verbal schemes in order to attach meaning to the information.

However, a common problem that can be found in ontology systems is the possibility of unexpected or unknown situations due to limitations in acquisition, processing, and interpretation of data.

Quality is recognized as a relevant performance issue operating processes, as well as in situations of understanding process for decision-making, since bad identification of the complaint leads to several failures in both resource allocation and tactics definition to attend such 190 complaints.

In order to contribute to SAW reported by emergency calls, this work constitutes a step of evaluation soon after a robbery complaint reported by the victim. Hence, the overall objective is to define a methodology for evaluation and representation of quality data from robbery complaints reported to the military police of the State of São Paulo in Brazil. Quality functions will be defined for detection of possible problems according to specific dimensions such as completeness and timeliness.

This methodology aims to provide support for the development of military decision-making support systems by means of objective and subjective evaluation quality metrics, as well as an ontology for the representation of mastering the knowledge. To illustrate the applicability of the methodology, it will be conducted a case study in which the victim is robbed and makes an emergency call. Data quality assessment functions are applied to this call, which in turn will be molded according to an ontology.

2 Data Quality for the Evaluation of Situations for General Decision-Making Systems

Data quality can be defined as one crucial factor in decision-making systems. Imperfect information, which do not truly describe real world situations, reduce the

effectiveness of the systems, contribute negatively to the mental model formation and, consequently, undermine the SAW process.

According to the literature, there is not a defined pattern to data quality in decision-making systems. Requirements are divided into dimensions or metrics, and their applications are highly domain dependent, whereas the application defines their respective meanings according to objectives, tasks and associated decisions. For robbery domain, this work methodology addresses quality ratings about completeness and timeliness. Approaches, descriptions, and different data quality perspectives are described below.

O'Brien [1] defines data quality dimensions required for information systems in three main dimensions: content, time, and shape. Among the quality attributes there are readiness, acceptance, frequency, period, accuracy, relevance, completeness, conciseness, breadth, performance, clarity, detail, order, presentation, and media.

Wang and Strong [2] categorize quality dimension attributes in four main classes described as follows:

Table 1. Methodologies for quality evaluation

Reference	Definition	Methodology
Yang W. et al. [3]	Methodology to help organizations to assess the data quality state in as well as monitor such quality improvements over time	1- Performance model of products and services by dividing a fixed set of criteria for data quality in four classes. 2- 65 survey evaluation items developed in accordance with the model. 3- Two analysis techniques: the first technique compares the quality of information according to the organization standard of best practices; the second technique measures the distance between different stakeholder's assessments about a system.
Bobrowski et al. [4]	Methodology for organizational data assessment	1- Definition of a list of criteria for information quality divided between direct and indirect assessment criteria. 2- Scores for indirect quality assessment criteria are computed by direct evaluation criteria, such direct criteria are defined by means of questionnaires.
Kim et al. [5]	Study to measure the attention engagement	1- The engagement attention of the participants is measured while they watch a particular movie. 2- Quality evaluation is carried out by means of a questionnaire, electroencephalogram and statistical analysis.
Batini et al. [6]	Methodology set out in order to assess and improve all kinds of data management in an organization, be it structured, semi-structured,and unstructured data	1- Reconstruction of relevant knowledge about organizational units, processes, resources and conceptual entities involved in the organization through interviews and questionnaires. 2- Quantitative evaluation of quality problems. 3- Improvements and enhancements.

Intrinsic data quality implies guaranteeing credibility and reputation to data, among the attributes there are credibility, reputation, accuracy and objectivity.

Contextual data quality is comprised by attributes that should be considered and evaluated according to the context of the task to be performed, having as attributes: relevance, timing, completeness, etc.

As for the quality of representation, the attributes are defined according to the given format-related aspects (such as conciseness and representation), and the meaning in the understanding and interpretation of such data. Finally, the authors classify individual accessibility-related attributes.

It is important to note that in addition to the works cited, most works analyzed presented in their methodology a subjective analysis step performed by experienced users in the field by means of questionnaires, interviews or surveys as shown in Table 1.

Among the related methodologies in the literature that evaluate data quality and relate to military decision-making context, it is noted that most of them are applied after the crime event, during the report recording process in order to prevent spread low-quality reports in the system. [8] [9]

It is evident that there are several applications of quality dimensions in both military and other application areas

3 A Methodology for the Assessment and Representation of the Quality of Data from Robbery Events

In order to enhance SAW, this work aims at promoting two evaluations of the information management system operator of emergency call service of the military

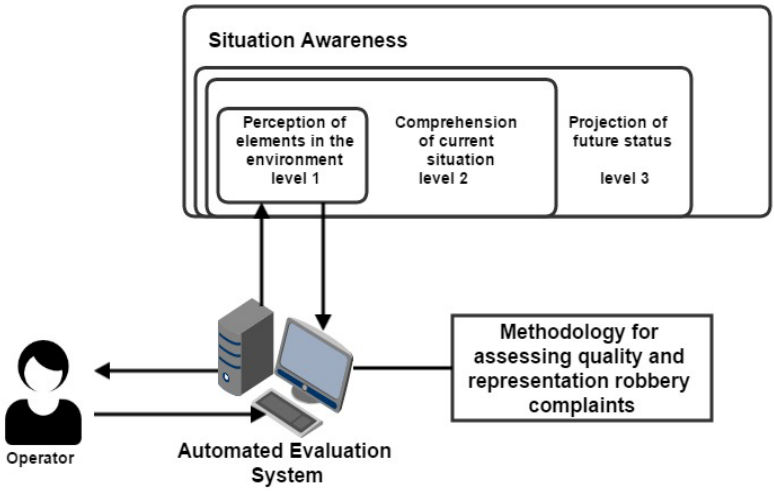

Fig. 1. Suggested model for methodology application focused on SAW's level 1

police. As well as contributing to accomplish first SAW level as exemplified in Figure 1, in which the system developed based on the methodology helps the perception of elements and its importance in the scenario. The methodology was defined considering decision-making systems, in which the operator can request more information by using a multiple data sources combination (data fusion).

To fulfill this objective, two methods are employed, classified as subjective evaluation, to be carried out by users, and objective evaluation, carried out according to a historical database supplied by values daily held in the system. The subjective evaluation performed by the operator consists of a list of criteria (the attributes defined for a robbery complaint) in which the scores are assigned to each item in the list.

Figure 2 presents an overview of the methodology addressing the two approaches (subjective and objective), to be described in more detail in this section. Initially, by system interface, the operator performs the score definition for the robbery complaint by using a list of values, and this process must be performed before any service task. Such knowledge is stored in database and the results are later represented by means of an ontology.

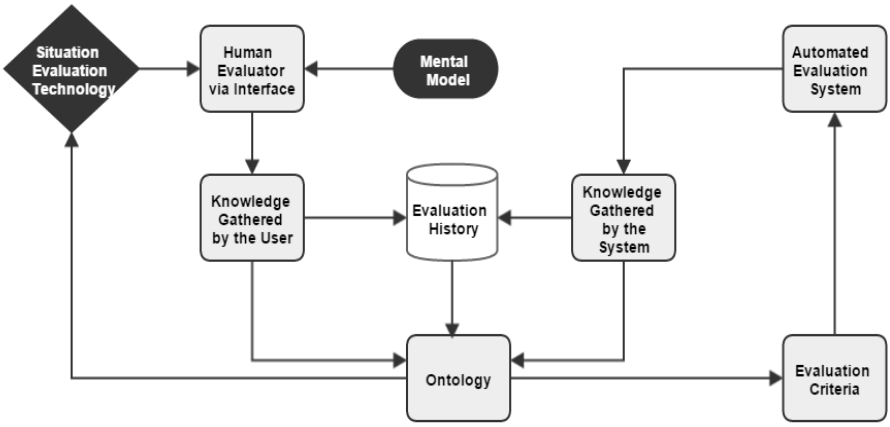

Fig. 2. Methodology's flowchart of evaluation and quality representation

When the system receives a new complaint, the real-time evaluation begins, and based on a model set for robbery complaints, it is carried out the completeness comparison where each item in the complaint receives the previously set score. As a result, it is obtained information about how the complaint is completed; and represented to the user by informing about missing items, if the operator deems necessary requests more information that will pass again through the process of comparison and assignment of scores.

Therefore, the proposed methodology is divided into three conceptual steps, which follows: (1) data quality requirements elicitation for military domain, (2) quantitative metrics and functions definition for requirements classification, and (3) knowledge representation of the domain through an ontology.

3.1 Requirements Elicitation for Military Domain

The requirements needed for the understanding of emergency calls were defined with the help of the military police of São Paulo. Initially, in order to perform the gathering of requirements, essential premises were defined to facilitate the decision-making process in the step of information visualization in the system.

In order to define how a complaint reported is complete, this phase presents the definition of a model of players and attributes for a robbery complaint. With the information obtained by GDTA (Table 2) it was modeled an attribute tree according to requirements in a robbery complaint. From the requirements, it was defined the components present in an event of robbery: the victim, the perpetrator, the stolen object, and the place and time of the event. The following is a description of each attribute:

- Criminal and Victim, who have similar attributes as individuals: clothing, characteristics, ornaments, and respective descriptions;
    - The perpetrator, in turn, has specific components such as current location, which consists of setting directions, driving, street or even neighborhood where the criminal headed out after the act.
- Object defines characteristics of the stolen object such as color, brand, size, and model. There is also an extension named Vehicle with specific features such as license plate and year in case such information is provided.
- Event spot, component provided with some type-specification (house, land, apartment, square) and with the information related to the address such as street, neighborhood, etc.

The Attributes Tree plays an important role in the next steps, since quality evaluation and knowledge representation of the domain take place by means of the objects and attributes defined.

Table 2. Robbery requirements set thru GDTA

SAW's level 1 Information to stimulate operator awareness
- Victim condition
- Potential start and end of event
- Event spot or surroundings
- Number of suspects
- Object in the suspect's possession
- Perpetrator's physical attributes
- Stolen object and its characteristics
- Quality score and data source (timeliness and completeness)

3.2 Quantitative Metrics Elicitation for Quality Evaluation

Ratings are applied according to timeliness and completeness data quality dimensions, with the purpose of distinguishing ambiguity about the real time of robbery events and the complaint itself. Temporal data identification is utterly important, since in certain cases the complaint of robbery can be notified in real time during the event, seconds, minutes or hours later. The identification of these temporal aspects assists in the definition of the plan of action to be followed, as well as which category of public safety will be responsible to meet the emergency (civil police, military, etc.).

Temporal evaluation is carried out taking into account three factors: the time at which the event takes place, the time at which the complaint is made, and the time the complaint in the system is recorded. Since a complete perception of the situation elements helps the operators understand the elements, the evaluation of the variation of these three factors can contribute to the decision-making, positively if based on actual data, or negatively if temporal data are incomplete or misinterpreted.

To capture knowledge from the user, it was defined a list composed by objects and attributes of a complaint in which the user assigns percentage values of how important he designates each attribute. The values are stored in a database, providing basis for complaints for the current day's tasks. It is proposed a daily application, because the user's mental model is improved according to their experience when performing daily activities. The daily application of values assignment may vary depending on these factors, thus attributing such flexibility to the ratings.

By means of the system, the user assigns values to the listed attributes. This knowledge is stored in the user-generated ratings file and used for the evaluation of quality when the complaint is logged. The evaluation process works as follows; the system receives a string containing the denunciation of the robbery. Considering a database of words identified for each attribute of the complaint, it is performed, the quality assessment regarding completeness, in which a comparison is made in the search of existing items in the complaint based on the defined model and based on the form of values defined for the day's activities. Then the existing items in the complaint receive such percentages.

In an objective evaluation, the scores are assigned according to an average of forms saved in the records and assigned to the items present in the complaint according to the model. The evaluation process is carried out in a similar manner to the subjective evaluation, where the complaint is received through a string. The completeness assessment is applied by checking the occurrence of the attributes and their corresponding values. The final representation results in the evaluated attributes and their respective percentages of quality.

3.3 Knowledge Representation

Matheus et al. [7] discuss the development of an ontology to improve SAW, which would be a basis to define ontology in the present work, since it aims to represent the objects and the relationship among them, besides their respective evolutions over time.

SAW's core ontology defines attributes and relationships where *SituationObject* are entities in a situation that may have characteristics (attributes) and can participate in relationships. The attributes set values of specific features of objects such as weight or color. *PhysicalObject* is a special type of *SituationObject*, which has specific values such as volume, position and speed. Relationships define the values of relationships (only true values were considered) between defined sets of *SituationObject*. They are usually derived from the system, but also from the possibility of being reported by external observations.

Attributes values and relationships are defined by a common class called *Propertyvalue*. This class provides a function of time-dependent value over a specific range. The time interval is set in relation to a *StartEvent* and possibly an *EndEvents* associated with specific *NotifyEvent.*

Since ontologies for specific domains start from top ontologies in order to contribute to the improvement of SAW, the ontology of the work was developed based on the ontology for SAW, based on requirements defined by the robbery complaints tree. The ontology was developed starting from the premise where the control system of emergency call service processes data from heterogeneous sources by data fusion in order to enhance the completeness of a complaint notified, such as social networks, cameras, videos, etc.

4 Case Study

This case study presents a situation in which SAW is a paramount factor for decision-making about resources allocation of the military police of the State of São Paulo (PMESP). Given the large amount of complaints of criminal activities aimed at PMESP, the main objective of this study is the identification, confirmation and understanding of occurrences, aiming at the reduction of meeting the same for the allocation of military resources efficiently. Additionally, we seek to identify and understand contexts associated with the situation, such as location, offender, stolen object, presence of weapons and victims. This case study discusses specifically the situation of robbery. The occurrence is initially reported via phone (190), and then applied the functions of quality involving the completeness and temporal aspects in the data about the events.

Case description: At 7 p.m., after returning from work driving his car in the southern region of Sao Paulo, John was approached by two individuals on a motorbike, while waiting to cross a traffic light in Vila Mariana. John was threatened, beaten and forced to leave the vehicle. One of the criminals dropped the bike and assumed the direction of the car. Then both fled towards western region. It was rush hour, so several people witnessed the action. One of the witnesses called 190. In the emergency call center of the PM, constable Campos answers the witness and asks initial information of the event. It was a hectic day. Several instances of robbery were reported that night. Not all were met with success. Several false or inconclusive calls soon made constable Campos unmotivated. Thus, the witness, very agitated and tense describes what happened:

(Phone call*): "Good evening! A carjacking has just happened at Domingos Setti with Luis Vives streets. Two guys riding bikes pointed a gun to the driver of a black Mercedes and made him leave the car without taking anything. They two fled speeding toward Klabin subway station. "*

The methodology aims to contemplate control systems of emergency calls, where in order to perform a completeness assessment, the operator performs fusion with heterogeneous data sources in order to increase the index of completeness of such complaint. Thus, the projection of the time evolution of completeness and its contents plays an important role in the process of SAW operator.

Whereas the assignment list of values for the attributes of the complaint was held, according to the requirements defined by means of GDTA, when the complaint is received it is applied the assessment method for completeness in order to quantitatively define if the emergency call was complete. In this manner, it is at this time that completeness is applied by returning a score related to the call. At the same time, it is carried out the time evaluation of the complaint by means of the information available to the complainants seeking to distinguish the time of the robbery, the registry made in the system, and current time.

Then the components are correlated by means of the ontology developed for representation of complaints of robbery based on the ontology for SAW. Critical classes to improve situational awareness were adapted to the specific scenario such as *ObjetodaSituação*, which consists in a set of objects related to the robbery such as *ElementosRoubo* (local, criminal, victim, stolen property, etc.). Event Notification is received by the system, and Attributes record the evolution of the attributes requested in case of new information required by the operator.

Notifications happen by means of the service system of emergency calls, where the first notification always matches the complaint made by the victim. The robbery is composed of a set of elements, has specific attributes such as description, event time and matches. The first notification is always the notification of a complaint, in which the urgency of a robbery near notifications can be related to previous complaint or not. This relationship is performed by evaluating the completeness of complaints received.

Through the fusion of data held in the system aiming at further information for a particular robbery, a new notification is handled by the *ElementosRoubo* class that manages elements of the robbery be it the victim, the perpetrator, the object or the location. Characteristics and specific attributes of the elements are defined in the class attributes. Subsequent notifications complement the robbery complaint related data to update the attributes of elements.

An important point is that even in the face of factors that can influence the operator to perform the definition of values such as tension, for example, the individual evaluation performed by the system based on the already set completeness scores helps minimize negative factors of subjective evaluation. Finally, the ontology represents the results, in which the operator, based on their experience and prior mental model, can achieve SAW to meet emergency calls.

5 Conclusion

This work presented a data quality evaluation methodology to enhance the first level of situation awareness to support the perception and understanding of robbery events and promote better military responses. Such methodology was divided in three steps. At first, a requirement elicitation was held thru the GDTA methodology in which an attribute tree was stablished with main robbery objects and its attributes, being: criminal, victim, event spot, time and stolen object.

Then, two completeness assessment metrics was set divided into subjective and objective evaluation. By means of a questionnaire composed by the robbery objects and its attributes, the user operator defines quality scores fomenting objective evaluation to be performed in real-time of robbery event by the system. Finally, a domain ontology based on saw core ontology was set to provide semantic meaning and relationships between objects and attributes while performing data quality evaluation.

Hence, it is concluded that the assessments of the data contained in each emergency call provide a full perception of the entities of an event and the necessary information about the real-time situation. The knowledge generated may assist the development of systems that demand SAW, since the evaluation of quality tends to improve the representation of both present and absent complaint information. Since the steps of the methodology were set to focus on the perception of the elements, it is believed that the goal has been accomplished successfully by identifying elements present in reports of robbery events by highlighting them and setting scores of completeness.

References

1. O'Brien, J.: Sistemas de Informação e as Decisões Gerenciais na Era da Internet, 2nd edn. São Paulo: Saraiva, vol. 2, p. 431 (2004)
2. Wang, R., Strong, D.: Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems, 5–33 (1996)
3. Yang, W., Strong, D.M., Kahn, B.K., Wang, R.Y.: AIMQ: a methodology for information quality assessment. Information & Management, 133–146 (2002)
4. Bobrowski, M., Marre, M., Yankelevich, D.: A Homogeneous Framework to Measure Data Quality. In: Fourth Conference on Information Quality, pp. 115–124 (1999)
5. Kim, W., Lee, S., Kim, Y., Choi, H.: Evaluating the Data Quality and the Uncertainty in Electroencephalogram Signals for a Neuromarketing Service which Computes Attentional Engagement, pp. 62–66 (2014)
6. De Amicis, F., Batini, C.: A Methodology for Data Quality Assessment on Financial Data. Studies in Communication Sciences, 1–12 (2004)
7. Matheus, C.J., Kokar, M.M., Baclawski, K.: A core ontology for situation awareness. In: Proceedings of the 6th International Conference on Information Fusion, FUSION 2003, vol. 1, pp. 545–552. IEEE Computer Society (2003)
8. Laudon, K.C.: Data quality and due process in large interorganizational record systems. Communications of the ACM (1986)
9. Bureau of Justice Statistics, U. S. Assessing completeness and accuracy of criminal history record systems: audit guide. Boureal of Justice Statistics. U.S. Deptartment. of Justice, Office of Justice Programs, Indiana, 65 (1992)