Developing a core ontology to improve military intelligence analysis

Valentina Dragos

Developing a core ontology to improve military intelligence analysis

Valentina Dragos

ONERA - The French Aerospace Lab
F- 91761 Palaiseau, FR
valentina.dragos@onera.fr

Abstract. In highly dynamic and heterogeneous environments, providing commanders with decision making support requires a through understanding of processes involved and the development of underlying knowledge models upon which reasoning mechanisms can be based. This paper presents the construction of ONTO-CIF, a formal ontology created to improve intelligence analysis. ONTO-CIF was developed by following a methodology based on textual documents, which allows us to accomplish a satisfactory accuracy level in terms of domain coverage while remaining on a manageable scale size. The paper also illustrates several semantic-based scenarios to support intelligence analysis, a central task of the military application field.

Keywords: Ontology, formal model, defence and intelligence

1 Introduction

Intelligence analysis is one of the key activities undertaken by intelligence staff, from lower echelons up to higher levels of the military chain of commands. In the military field, this activity supports decision making based on reasoned interpretation of the current situation, backed up with solid evidence from various intelligence sources.

With the increasing number of information sources, it becomes important to incorporate intelligence in a secure and reliable manner. While algorithms developed for object detection and tracking [4] or images analysis [3] offer good results, they prove to be limited as soon as we are handling soft data, such as image annotations or HUMINT [1] reports. Beyond large topical and linguistic variety, solutions for symbolic data treatment should take into account several particularities of the military field: differences between doctrines, various security levels and cultural differences. Those constraints give rise to a need for explicit representation of domain-knowledge to be used as a basis to develop enhanced treatments for intelligence analysis.

The goal of this work is to build a formal ontology for intelligence, providing a standard model of concepts and semantic relations. We also discuss funda-

[1] HUMan INTelligence: intelligence gathered by the means of human sources

mental questions underpinning ontology construction in the military field and we introduce several semantic-based scenarios to support intelligence analysis.

The outline of the paper is as follows: section 2 provides a brief overview of ontologies in the military field. The application context is presented in section 3, with emphasis on specific concepts and user needs. Section 4 details the construction of ONTO-CIF[2] while section 5 discusses technical bottlenecks of intelligence analysis and provides semantic-based solutions to overcome them. Section 6 briefly discusses critical aspects of ontology construction and the paper concludes by final remarks and perspectives for future work.

2 Ontologies for military applications

The use of ontology has attracted much attention within the military field, with the rise of new treatments allowing for the domain knowledge to be incorporated within different systems. Focus lies on developing ontologies to improve situation awareness, to facilitate information integration for military coalition and to provide decision support for the command and control process. This section presents a brief overview of those ontologies.

2.1 Ontologies for situation awareness

Situation awareness enables an agent to gather information about objects (vehicles, pedestrians, etc.) evolving in complex environments such as battlefields or urban areas.

One of the first models developed to support situation awareness is the SAW (Situation Awareness) ontology [11]. SAW models a *situation* as being composed of *entities*, interacting in order to achieve one or several *goals*. Entities are physical or abstract objects and their interactions are modeled by semantic relations. By using SAW it is impossible to identify the spatio-temporal context of actions as space and time concepts are not considered.

Coping with this limitations, CONON (CONtext ONtology) ontology, described in [19] offers a conceptual description for situation awareness having the *context* as central element. The context is an aggregation of four features: *local coordinates*, *persons* and *objects* along with their respective *activities*. Although not explicitly defined, a situation can be created by using a set of concepts. More details about ontologies for situation awareness are provided by [2].

2.2 Ontologies for military coalitions

Military coalitions are missions conducted jointly by several armies against common adversaries. A main issue is related to information exchange between coalition members and ontologies are used to offer a common understanding of ongoing actions and to improve data and information interoperability.

[2] ONTOlogy for information sCorIng and Fusion

An ontology of military coalitions was created by Dorion and colleagues, [6]. The model is based on the C2IEDM (Joint Consultation Command and Control Information Exchange Data Model), [14] and is was used to support the development of automatic procedures to process textual messages compliant with the standard T-OTH-GOLD (Over-the-Horizon Targetting GOLD)[3]. This ontology highlights five main concepts, corresponding to: *individuals*, *organizations*, *facilities*, *equipments* and *features*. Entity *types* (military vs.civilian), *affiliation* (organization membership), *status* (hostile vs. allied) and *geographical location* are also modeled.

2.3 Ontologies for command and control

Command and control is a critical process for successful military operations. It makes use of advanced information processing techniques in order to provide valuable information to feed the basic building blocks of decision support systems.

Developing a conceptual model to support this process is a challenging task: the outcome is required to have a large coverage, as each information piece can be relevant. Currently, two complementary approaches are adopted. The first one is proposed by [12] and it creates a single ontology by using automatic procedures able to translate entities of the JC3IEM model (Joint Command and Communication Information Exchange Data Model), [15] as ontological entities. The translation is carried out thanks to transformation rules established by domain experts. The outcome is a large-size ontology (7900 entities), having a good quality as it relies on a commonly accepted model. The second approach is adopted by [18], and it creates several ontologies, describing five sub-domains: *entities*, *areas of interest*, *convoys of vehicles*, *infrastructures*, *conditions* and *alerts*. When using those ontologies, a mapping phase is required in order to create bridges between them, which can be an expensive task.

2.4 Yet another ontology for military intelligence

This paper presents the development of a formal ontology for intelligence analysis and the work is part of efforts conducted to support the command and control process. Creating a new ontology is needed as currently existing models prove to be either too general to fit our envisioned task, or tailored for different applicative scenarios, and therefore inappropriate.

Our approach consist in identifying a particular task of the command and control process (intelligence analysis) and building a core ontology, highlighting only main concepts related to this task. The ontology is developed to have an appropriate specialization level to describe fundamental aspects of military intelligence, while remaining on a manageable scale, for ease of use.

[3] Multilateral Interoperability Programme Web Site: www.mip-site.org

3 Application context: intelligence analysis

Human operators receive intelligence in the form of summaries, briefings or more elaborated reports describing critical facts. Sometimes, interpretations or recommendations about actions to engage are appended to intelligence reports.

As there is a stream of incoming intelligence, it is not always easy for operators to determine whether an item is compromised or reinforced by new information. More over, information can be unreliable, inaccurate or old, and overlaps can appear as different sources provide the same information.

The goal of intelligence analysis is to create richer descriptions of events or actions and to provide a means to assess the accuracy of those descriptions. Supporting this task can be achieved by providing strategies for information fusion, on one hand, and techniques and algorithms for information evaluation, on the other. Therefore *information fusion* and *information evaluation* are the main concepts of intelligence analysis.

We adopt the definition of *information fusion* as provided by the international society in information fusion (ISIF[4]): *information fusion encompasses the theory, techniques and tools conceived and employed for exploiting the synergy in the information acquired from multiple sources (sensor, databases, information gathered by human, etc.) such that the resulting decision or action is in some sense better than it would be possible if any of these sources were used individually without such synergy exploitation.*

Information evaluation is considered with respect to doctrinal documents of military intelligence [16]. Thus, information items are assigned two attributes: the *reliability* of source and the *credibility* of information. *Source reliability* is the degree of confidence assigned to a source with respect to its capacity to provide valuable and accurate information. Information *credibility* measures the confidence assigned to information with respect to its veracity. Both source reliability and information credibility are dynamic measures.

From a practical standpoint, information fusion and evaluation are complementary user needs. Les us consider the following information items provided by two distinct sources:

(I_1) Civilian 4x4 vehicles parked on access road to MSR

(I_2) Military 4x4 vehicles parked on access road to MSR

If the user needs to be informed about the presence of a vehicle around MSR area, then information fusion can produce a new item of information such as:

(I_3) (Military or Civilian) 4x4 vehicle parked on access road to MSR

If the user needs to be informed about the presence of a civilian vehicle around MSR area, then I_1 and I_2 are contradictory items of information. Thus, information evaluation outputs low credibility values for each item and it is impossible to fulfil the user's request.

[4] http://www.isif.org/

4 ONTO-CIF: a core ontology for military intelligence

4.1 Knowledge representation and abstraction level

Ontologies are formal and explicit specification of a shared conceptualization, [9]. An ontology is a conceptual model highlighting domain-specific concepts and semantic relations. Concepts are formal representations of entities, while relations correspond to interactions or associations of entities. The subsomption relation creates a structured taxonomy of concepts, through which inheritance mechanisms can be applied. Ontologies also model axioms, as logical expressions which are always true. By making domain knowledge explicit, ontologies provide a good basis for reasoning mechanisms. They also offer a means to handle linguistic variety, as ontological entities are named by natural language terms.

Ontologies are classified in three main categories, according to the abstraction level of knowledge represented: upper ontologies model high-level knowledge, based on philosophical reflections; core ontologies describe structural domain concepts and relations while domain ontologies provide detailed models of concepts as defined and handled by domain experts.

With respect to the classification above, we define ONTO-CIF as a core ontology, created to support intelligence analysis. It provides a generic description of the application field and its enrichment is needed in order to fit specific tasks. By choosing this abstraction level our intention is to manage the trade-off between the domain coverage and its detailed description, and to facilitate the exploitation of ONTO-CIF.

4.2 Construction of ONTO-CIF

To create ONTO-CIF we followed a four phases construction process, see fig.1, described here after.

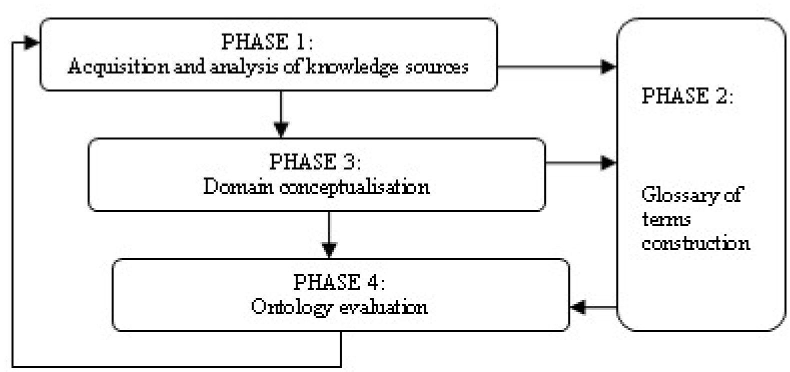

Fig. 1. Overview of ONTO-CIF construction

Phase 1: Acquisition and analysis of knowledge sources. This phase aims to create a collection of relevant knowledge sources to support ONTO-CIF development. For this work we used NATO standardization documents. By exploiting a collection of documents created by domain experts, we ensure the construction of a consensual representation shared by domain actors with respect to both conceptual and lexical points of view.

In the light of those sources, a set of five notions corresponding to *Event*, *Place*, *Organization*, *Person* and *Equipment* (the pentagram, see fig. 2) appears as the central element of our domain. Collected documents provide definitions

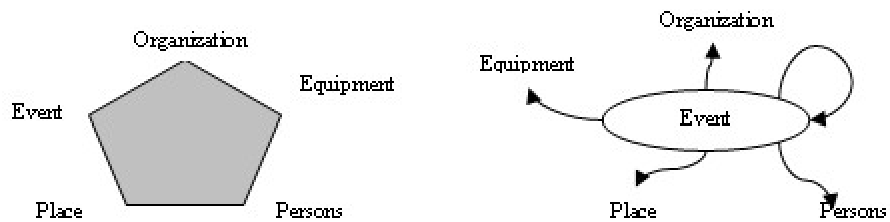

Fig. 2. Intelligence main concepts (the pentagram) and their associations

of pentagram notions (i.e. an equipment is defined as '*any item of materiel used to equip a person, organisation or place to fulfil its role*'), and descriptions of entity associations, as shown in fig. 2. Some documents also specify procedures to follow in order to ensure the quality of information exchanges.

The analysis of collected sources highlights several issues to take into account in our approach to model the military intelligence field. Hence, there are no standardised taxonomies for military intelligence, although the AIntP-3 resource [17] was developed to facilitate exchanges of military data. At a more general level, NATO maintains various resources, such as the APP6 glossary [5]. As a consequence, those different attempts to conceptualize the domain introduce alternative and sometimes inconsistent definitions of the same concept.

Moreover, the domain is modeled by considering different points of view: for instance, AIntP-3 provides functional descriptions of entities, while App6 is rather a collection of symbols used to mark military maps.

We can also mention difficulties related to the description of concepts referring to the same feature considered at different abstraction levels or the existence of fuzzy concepts, belonging intrinsically to several categories.

Phase 2: Glossary of terms construction. This phase consists on the selection of domain specific terms. The glossary is created by including relevant terms naming concepts, instances, attributes or concept associations, along with their synonyms. We restricted the construction of this glossary to the collection of

[5] http://www.nato.int/docu/stanag/aap006/aap6.htm

documents, therefore no external resource such as WordNet, [13] were used to enrich the set of synonyms. The glossary of terms if composed of 41 single-word or composed terms, see tab. 1. At this stage of the ontology construction, several

Table 1. Excerpt of the glossary of terms

Terms	Composed terms
Company	Social grouping
Flotilla	Terrorist attack
Civilian	Military chain of command

terms can refer to the same notion, and the same term can define both a concept and a relation.

Phase 3: Domain conceptualisation. The goal of this phase is to model domain concepts, to identify relations holding between them and to define formal axioms. The pentagram is the starting point of this phase, therefore we create a basic ontological structure composed of: *Person*, *Organisation*, *Equipment*, *Locations* and *Events*. Then we gradually enrich this basic structure, by using the glossary of terms. For instance, *Geographical area* and *Vehicle* are added as specific types of *Location*, and respectively *Equipment*.

If the pentagram highlights concepts corresponding to the most relevant domain entities, those categories prove to be insufficient as soon as we are interested in describing the status of entities or their structures. To cope with those limitations, we introduce 3 abstract concepts, describing respectively: status of entities, according to various points of view (ex. *Military* vs. *Civilian*, *Natural* vs. *Artificial*) their structure (ex. *Hierarchy*) and their goals (ex. *Transportation*, *Communication*, etc.). During this stage we have identified 58 concepts, clustered in 6 main categories, as shown in tab. 2. By taking into account temporal aspects,

Table 2. Excerpt of ONTO-CIF concepts

Category	Concepts
Entity	Organisation, Person
Event	Terrorist attack
Location	Place
Status	Civilian, Military
Entity-Aim	Goal
Structure	Hierarchy

we introduce *Entity* as the most general concept regrouping time-independent artefacts and *Event* as the most general concept corresponding to some action that takes place at a given place and time, for instance *Terrorist Attack*. Ontological relations are modeled by exploiting the set of associations identified during the previous phase. Those relations are: specializations, compositions and field-specific relations, see tab. 3.

Table 3. Excerpt of ONTO-CIF relations

Relation	Concepts related
Specialization	Organization, Social-Grouping
Composition	Organization, Person
Equivalence	Place, Installation

Axioms of ONTO-CIF define equivalent classes (*Installation* = *Place*), disjoint classes (*Military* vs. *Civilian*) and reversed roles (*function-Of(Function, Installation)* vs. *has-Function (Installation, Function)*).

Phase 4: Ontology evaluation is an aspect unaddressed in this paper. We assume that ONTO-CIF has a good domain coverage, as the construction exploits a collection of documents created by domain experts. Section 6 provides a discussion on particular issues related to the evaluation of ONTO-CIF.

4.3 Description of ONTO-CIF

ONTO-CIF was implemented by using Protégé [6] (fig. 3) and it is represented in OWL DL, a sub-language of OWL [5]. The model is composed of 54 concepts,

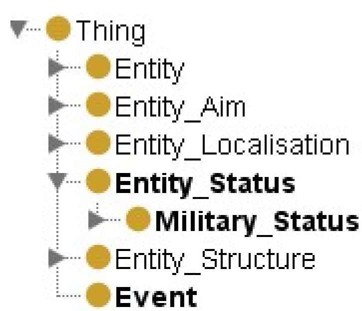

Fig. 3. ONTO-CIF : hierarchy of concepts

[6] http://protege.stanford.edu/

related by 58 relations and it highlits 4 pairs of equivalent classes, 6 pairs of disjoint classes and 4 pairs of reversed roles.

5 On the use of semantics to support intelligence analysis

Ontologies are artefacts modeling domain knowledge by taking into account both conceptual and linguistic levels. The conceptual level concerns the description of field entities, while the linguistic level is related to the use of natural language terms to name ontological entities. By making domain knowledge explicit, ontologies provide a good basis for further automatic reasoning mechanisms. By offering this two-fold description of domain knowledge, they also offer a solution to handle linguistic variety.

5.1 Technical bottlenecks of information fusion and evaluation

Technical bottlenecks of information fusion and evaluation are related to the identification of various information items connexions and their exploitation in order to infer new information or to complete the existing ones. More particularity, identifying redundant information allows us to avoid information overlap, while discovering contradictory information will highlight the existence of conflictual items.

From another perspective, the identification of correlated information reveals connections between information items and it is closely related to information enrichment, a process consisting in composing several information pieces in order to create a complete description of events or actions, and the identification of new information items. Here after we illustrate semantic-based solutions to overcome technical bottlenecks of information fusion and evaluation.

Identification of contradictory information highlights conflictual information provided by distinct sources or by the same source over the time. Disjoint concepts or reversed roles are central issues to retrieve contradictions, see fig. 4.

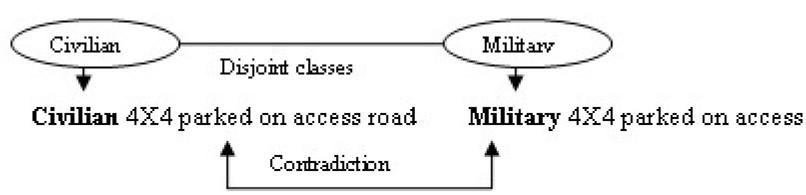

Fig. 4. Contradictory information

Identification of redundant information aims to retrieve identical items of information provided by distinct sources. Ontologies can support this task as they highlight the set of synonyms referring to the same concept, see fig. 5.

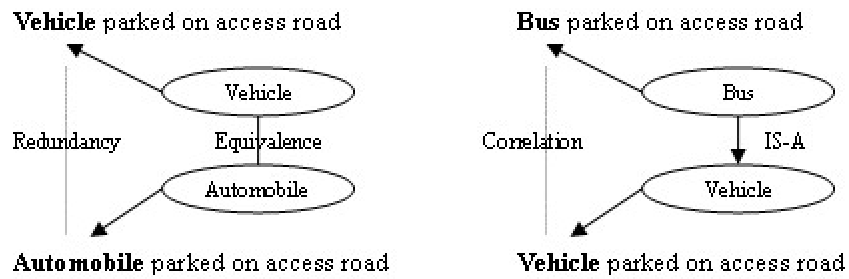

Fig. 5. Redundant and correlated information

In a similar way, ontologies appear as an appropriate tool to identify information correlations, as they provide explicit definitions of semantic relations, see fig. 5.

Information enrichment creates a richer information by combining various information pieces. This task is closely related to the previous one, as it links together information items according to their mutual correlations, see fig. 6.

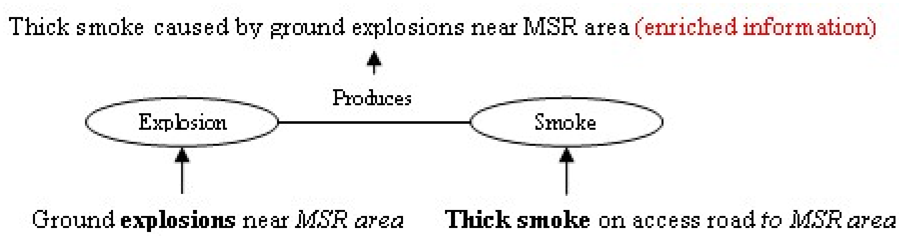

Fig. 6. Enrichment of information

For instance, it becomes possible to establish connexions between events occurring in the same place, at different moments of time, and therefore to describe the evolution of a particular scene.

New information discovery, see fig. 7 aims to identify previously unknown connections between information items. This can be achieved by performing logical inferences over the set of ontological entities.

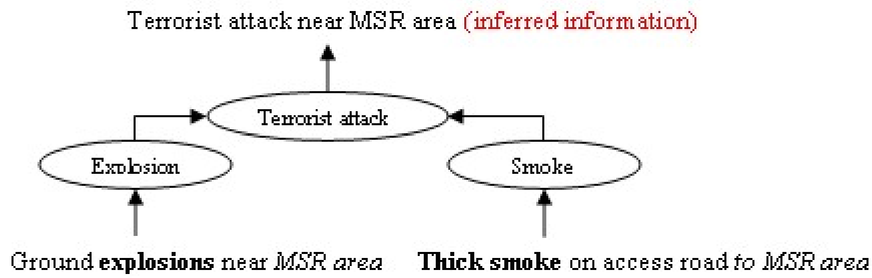

Fig. 7. Discovery of new information

5.2 Drawbacks and limitations of semantic-based solutions

Identifying information correlation is the main issue for intelligence analysis, along with the capacity to discover the type of those correlations such as confirmation or contradiction. For this reason, particular attention was paid to elements able to highlight those aspects during the ONTO-CIF construction process.

The main drawback of a semantic-based solution is related to its capacity to provide a *closed-world* model of the application field. If we consider the identification of correlated information, this is tributary to the set of relations already known and modeled. As the domain knowledge dynamically chances, the model should be adapted in order to keep an up-to-day field description.

Another limitation when using exclusively semantic-based solutions occurs as linguistic phenomena are not taken into account. For instance, contradictory information are also highlighted by negations, which are ignored when performing an ontology-based analysis. A more complex approach, using jointly natural language techniques and domain knowledge is suitable to offer a complete frame for enhanced intelligence analysis.

6 Discussion on model construction and evaluation

Difficulties of ONTO-CIF modeling are related to its manual construction.This approach was adopted as only a reduced number of textual documents were available to create the ontology. The outcome coverage is also a critical aspect. We consider the ontology as having a satisfactory domain coverage, as the sources used for knowledge acquisition are validated by domain experts and widely exploited within the application field.

Evaluation of ONTO-CIF is an aspect unaddressed in this paper. Evaluating ONTO-CIF give rise to particular issues, as it is a core ontology, whose practical use requires a previous enrichment process. As a future work, we envision to propose an evaluation protocol, taking into account the cost (understood as a measure of effort) associated to the adaptation of ONTO-CIF to different applicative scenarios.

7 Conclusion

Numerous research efforts have been made towards the formalization and exploitation of ontologies in support of military applications, leading to new approaches of information processing. This paper describes the construction of a core ontology for military intelligence and this work offered us the opportunity to broaden our comprehension of ontologies as an appropriate tool to handle technical bottlenecks of intelligence analysis.

The next step is the development of applications based on ONTO-CIF and our envisioned work concerns the adaptation of the ontology in order to implement a semantic-based approach to analyze HUMINT messages, [7]. First, ONTO-CIF will be enriched in order to take into account particularities of data to be analyzed. A more specific ontology will be obtained tailored to fit the task requirements. Then, this outcome will be used to define semantic similarity of HUMINT messages, offering this way a semantic-based application framework.

References

1. Baader, F., Calvanese, D., McGuinness, D., Nardi, D. and Patel-Schneider, P.-F.: The Description Logic Handbook : Theory, Implementation, and Applications. Cambridge University Press, 2003.
2. Baumgartner, N., Retschitzegger, W.: A Survey of Upper Ontologies for Situation Awareness. In Proceedings of the International Conference on Knowledge Sharing and Collaborative Engineering (KSCE '06), USA, 2006.
3. Bendjebbour, A., Delignon, Y., Fouque, L., Samson, V., Pieczynski, W.: Multisensor image segmentation using Dempster-Shafer fusion in Markov fields context, Geoscience and Remote Sensing. EEE Transactions on , Volume: 39 , Issue: 8 Digital Object Identifier: 10.1109/36.942557, Page(s): 1789 - 1798
4. Blasch, E., Lang H.: Data association through fusion of target track and identification sets. In Proceedings of the Third International Conference on Information Fusion, Canada, 2000.
5. Dean, M. and Schreiber, G.: Owl web ontology language reference. Technical report, W3C Proposed Recommendation, 2004.
6. Dorion, E., Matheus, C. J., et Kokar, M.M.: Towards a formal ontology for military coalitions operations. In Proceedings of the 10th International Command and Control Research and Technology Symposium, USA, 2005.
7. Dragos, V. : Shallow semantic analysis to estimate HUMINT correlation, In Proceedings of the 15th International Conference on Information Fusion, Singapore, 2012.
8. Fernandez, M. , Gomez-Perez, A. et Juristo, N. : METHONTOLOGY: from Ontological Art towards Ontological Engineering. Proceedings of the AAA Spring Symposium Series on Ontological Engineering, USA, 1997.
9. Gruber, T. : A translation approach to portable ontology specifications. Knowledge Acquisition, pages 199-220, 1993.

10. McGuiness, D.L.: Ontologies for information fusion. In Proceedings of the Sixth International Conference on Information Fusion, pp. 650-656, Australia, 2003.
11. Matheus, C. J., Kokar, M.M. et Baclawski, K.: A Core Ontology for situation Awareness. In Proceedings of the Sixth International Conference on Information Fusion, Australia, 2003.
12. Matheus, C., Ulicny, B.: On the Automatic Generation of an OWL Ontology based on the Joint C3 Information Exchange Data Model. In proceedings on the 12th International Command and Control Research and Technology Symposium, 2007.
13. Miller, G.: Wordnet : A lexical database for english. CACM, 38(11):39-41, 1995.
14. MIP: Multilateral interoperability programme, overview of the c2 information exchange data model (C2IEDM), 2003.
15. MIP: Joint C3 Information Exchange Data Model (JC3IEDM Main) 2005.
16. NATO: Annex to STANAG 2022 (Edition 8). Information handling services, 1997.
17. NATO: The military intelligence data exchange standard - AIntP-3(B), 2005.
18. Obrst, L., Stoutenburg, S., McCandless, D., Nichols, D., Franklin, P., Prausa, M., Sward, R.: Ontologies for Rapid Integration of Heterogeneous Data for Command, Control and Intelligence. Chapter in: Obrst, L., Janssen, T., Ceusters, W. (eds.) Ontologies and Semantic Technologies for the Intelligence Community, pp. 71-89. IOS Press, Amsterdam, The Netherlands. IOS Press book series: Volume 213 Frontiers in Artificial Intelligence and Applications, 2010.
19. Wang, X.H., Zhang, D.Q., Gu, T. and Pung, H.K.: Ontology based context modeling and reasoning using OWL, Proceedings of the 2nd IEEE Annual Conference on Pervasive Computing and Communication Workshops, USA, 2004.