16$^{\text{th}}$ICCRTS

Collective C2 in Multinational Civil-Military Operations

"Towards Ontology Matching Suitable for Information Integration in Time-Critical Situations"

Topic:Information and Knowledge Exploitation

Bjørn Jervell Hansen

Norwegian Defence Research Establishment (FFI)

P o Box 25; NO-2027 Kjeller, Norway

+47 63 80 70 88

bjorn-jervell.hansen@ffi.no

Towards Ontology Matching Suitable for Information Integration in Time-Critical Situations

Abstract

Information handling is important when conducting operations according to NATO Network Enabled Capability. Ontologies - explicit formal models - are being put forward as a promising component in solving challenges in this area.

The predominant approach utilising ontologies is to base solutions on several medium sized or small ontologies. However, when solving information handling challenges utilising several ontologies, there will be a need to link them - to perform *ontology matching*.

Ontology matching is typically being conducted in a semi-automated manner where either (a) a tool presents a list of possible matches for a human to pick the appropriate ones, or (b) an automatic decision is being made based on a score threshold set by a human.

In a time-critical situation, like for example a decision maker needing to utilise an appearing information source, such semi-automatic approaches are either not fast enough (case (a) above), or does not guarantee an ontology match of sufficient quality (case (b) above).

In this paper, ideas for an approach utilising deductive ontology matching are presented. Such an approach can offer ontology matching suitable for time-critical situations, fulfilling the timeliness and quality requirements outlined above.

1 Introduction

One of the basic tenets of NATO Network Enabled Capability (NNEC), is improved information sharing among military units in order to enhance information quality and, in turn, shared situational awareness. This is in turn anticipated to be an important contributor to build the decision superiority that in the end will lead to increased mission effectiveness (Buckman 2005).

For this vision to be fulfilled, the information from the various sources needs to be integrated by the information systems. This is not a straight-forward task, as the different sources often deliver their information using different formats and models, making it difficult to process the received information.

One particular issue regarding information integration in NNEC, and the issue in focus in this paper, is how to perform information integration when the involved parties have limited time to solve their information integration needs. These situations are in this paper called time-critical, and an example can be a decision maker needing to fetch critical information from an unmanned aerial vehicle (UAV) that happens to be in the same area but will not remain there for long.

In a dynamic information systems environment where information sources can suddenly appear, as portrayed in the NNEC vision, there is a great need for systems able to integrate the information as soon as it becomes available. Ontology-based information integration has in recent years been put forward as a promising solution to this challenge (de Bruijn 2004, Hansen 2008).

Ontologies – formal, computer-processable specifications of a domain vocabulary of concepts and axioms relating these concepts – are models of a particular domain, and the current trend in ontology modelling is to focus rather on small and medium-sized ontologies with manageable scope and size instead of elaborate large ontologies designed to cover everything (Hitzler et al. 2009, chapter 8.4), (Stuckenschmidt et al. 2009). Ontology-based information integration solutions thus have to deal with several ontologies that needs to be linked.

The process of linking ontologies is called ontology matching (Euzenat & Shvaiko 2007), and it is typically being conducted in a semi-automated manner where a tool presents a list of possible matches for a human to pick the appropriate ones. When a higher level of automation is needed, a tool usually makes the decisions based on a score threshold set by a human.

In a time-critical situation, a semi-automatic solution like the one outlined above will not be fast enough. Thus a higher level of automation is needed. A threshold-based automated solution can mend this problem, but such solutions cannot guarantee the quality of the resulting ontology links. They are thus not good enough in a military scenario where the appearance of erroneous links will not be accepted.

As a solution to this problem, this paper highlights *deductive ontology matching* - an ontology matching regime where the use of logical deduction provides automated ontology matching that guarantees the quality of the matching results provided that the input to the process is correct.

The main contribution of this paper is proposing the use of deductive ontology matching to support information integration in time-critical situations. An outline of a possible way to organise the ontology handling in order to perform this brand of ontology matching is also presented.

The paper is organised as follows: in Section 2, the role of semantic technologies in general, and ontologies in particular, in performing information integration is presented. Section 3 then introduces ontology matching, while Section 4 presents our proposal of using deductive ontology matching to perform ontology matching for information integration in time-critical situations. Section 5 presents other work related to this proposal, while Section 6 summarises the paper and outlines what further work is needed in order to establish the viability of the proposed approach.

2 Semantic Technologies as a Contributor to Information Integration Solutions in NNEC

Semantic technologies is a family of information technologies that utilise formal models. One of the main attributes of this technology family is that it can contribute to build flexible information handling systems – a very interesting property in the dynamic information systems environment foreseen to be the result of conducting operations according to NNEC.

What makes semantic technologies different from traditional information technologies is that it represents the meaning (semantics) of the data explicitly in semantic models, which are kept separately from the data itself and the program logic. The semantic models are called ontologies and are computer processable, making it possible for computers to share and work with the semantics with minimal, and in some cases no, human intervention. The way the ontologies are linked to the software also means that they can be exchanged at run-time, providing the resulting software with a flexibility unattainable using traditional information technologies.

Ontologies are formal, computer-processable specifications of a domain vocabulary of concepts and axioms relating these concepts, and they represent a core technology in semantic technologies. Ontologies provide the flexibility that makes semantic technologies interesting as an ingredient in building information systems for NNEC.

Semantic technologies are being put forward as an ingredient in information integration solutions (de Bruijn 2004, Hansen 2008). Due to the ontologies' key role in semantic technologies, such solutions are often referred to as ontology-based information integration solutions.

The current trend in ontology modelling is modularisation (Hitzler et al. 2009, chapter 8.4), (Stuckenschmidt et al. 2009) – the making of ontologies with manageable scope and size as opposed to elaborate large ontologies designed to cover everything. Ontology-based information integration solutions thus have to handle several ontologies that needs to be linked. The process of linking ontologies is called ontology matching, and will be presented in the next section.

3 Ontology Matching

This paper follows the ontology matching definition from Euzenat & Shvaiko 2007:

the process of finding relationships or correspondences between entities of different ontologies.

Further, Euzenat & Shvaiko defines the matching process as follows:

The matching process can be seen as a function f which, from a pair of ontologies to match o and o', an input alignment A, a set of parameters p and a set of oracles and resources r, returns an alignment A' between these ontologies: $A' = f(o, o', A, p, r)$.

This process is visualised in Figure 1.

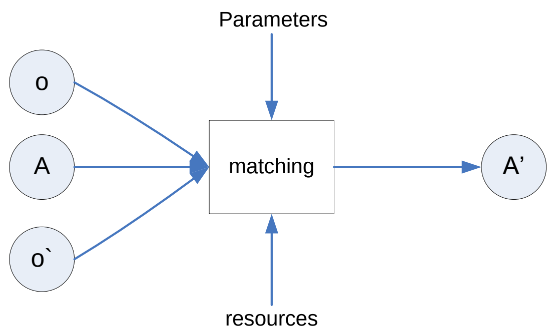

Figure 1: A schematic representation of the ontology matching process, taken from Euzenat & Shvaiko 2007

An alignment is further defined as:

a set of correspondences between two or more ontologies. The alignment is the output of the matching process.

Be aware that the term ontology alignment can also be used as more or less a synonym for ontology matching, like e.g. in Ehrig 2007.

Ontology matching can be performed manually by an expert examining the ontologies to be matched and deciding which concepts are related. This can provide a high-quality alignment, but is a time-consuming effort. In order to develop tools to assist users performing manual ontology matching, there has been developed a wide range of ontology matching algorithms. In Euzenat & Shvaiko 2007, these algorithms are classified in the following way:

- Name-based techniques (Comparing the string of the entity labels)
- Structure-based techniques (Comparing the structure of the ontologies)
- Extensional techniques (Taking the data adhering to the different ontologies into consideration)
- Semantic-based techniques (Deductive techniques using model-theoretic semantics)

In order to get an optimal alignment, different algorithms, often belonging to different classes in this classification, are normally used together.

An ontology matching algorithm often outputs a score on each possible match between the ontologies. An ontology matching assistance tool can then provide the user with a ranked list of possible matches, as illustrated in Figure 2. This list can then be taken into consideration when the user decides on the final alignment. Should a higher degree of automation be desired, this is normally attained by letting the user provide a score threshold above which the system accepts the correspondence as true. In the example in Figure 2, a threshold of 0.7 will for example mean that the correspondences `Source` ↔ `Observer` and `Track` ↔ `Observer` are approved, while correspondences `Source` ↔ `ObservationObject` and `Track` ↔ `ObservationObject` are not.

Alignments			
Ontology 1 concept	**Relation**	**Score**	**Ontology 2 concept**
#Source	=	0.876	#Observer
#Track	=	0.72	#Observer
#Source	=	0.56	#ObservationObject
#Track	=	0.2	#ObservationObject

Figure 2: An alignment score example

The latter approach is faster than approaches with a human in the loop, but has the problem that the quality of the resulting alignment is not guaranteed. Solving this problem while keeping the speed of which ontology matching is performed, is the topic of the next section.

4 Ontology Matching for Time-Critical Situations

For ontology matching to support ontology-based information integration in time-critical situations, it has to fulfil the two following requirements:

1. It has to be fast

2. It has to be reliable

Regarding 1, in the situations considered in this paper the information integration parties have limited time to perform the integration. Methods requiring a human user in the loop, as described in Section 3, will most likely not be fast enough.

Following the threshold-based automation strategy outlined in Section 3 can be a solution to requirement 1, but fails on requirement 2: with a threshold-based automatic decision there is no guarantees that there will not be erroneous alignments. Erroneous alignments can lead to poor integration, which in turn can lead to errors in the operational pictures. This can ultimately lead to situations resulting in loss of lives, thus erroneous alignments will not be tolerated.

One particular brand of ontology matching, however, has the potential to fulfil these two requirements: Deductive ontology matching. Deductive ontology matching techniques belong to the class of semantic-based techniques, and are based on deductive reasoning. When performing ontology matching using a deductive method, the correctness of the resulting alignments are guaranteed, provided that the inputs to the process, the premises, are correct. Ontology matching is, however, in essence an inductive task (Euzenat & Shvaiko 2007, p. 110), thus in order to perform deductive ontology matching a preprocessing phase providing the premises for the subsequent matching process is needed.

In this paper, the proposed premises needed in the deductive ontology matching process, are stored, high-quality alignments available to the ontology matching system. These alignments will be the results of user-guided ontology matching processes, assuring that they are of the needed quality.

The proposed process is shown in Figure 3:

1. The in-theatre decision maker in need of performing information integration fetches the ontologies representing the information sources in question

2. An alignment store consisting of certified alignments from previous ontology matching sessions is queried for all alignments which concerns a concept of either of the fetched ontologies

3. The ontologies, enhanced with the alignments, is processed with a reasoner in order to deduce alignments between the relevant ontologies.

In the easiest cases, alignments between the ontologies may already exist in the alignment store. However, an example as shown in Figure 4 is more suited to illustrate the possibilities in the approach: Two common operational picture (COP)

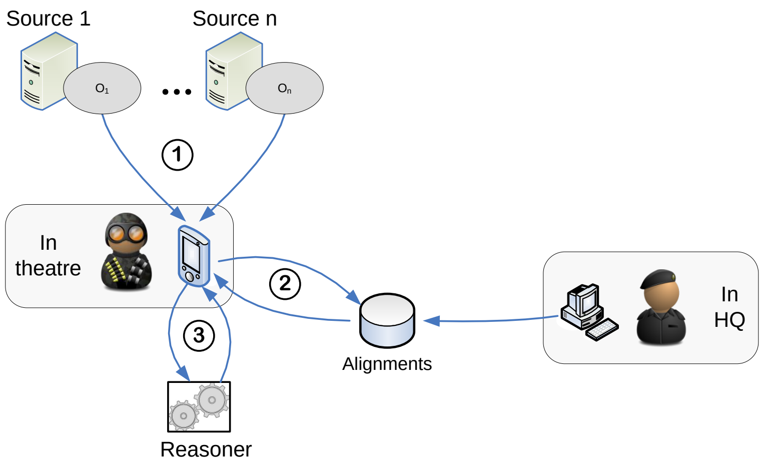

Figure 3: Proposed ontology matching architecture

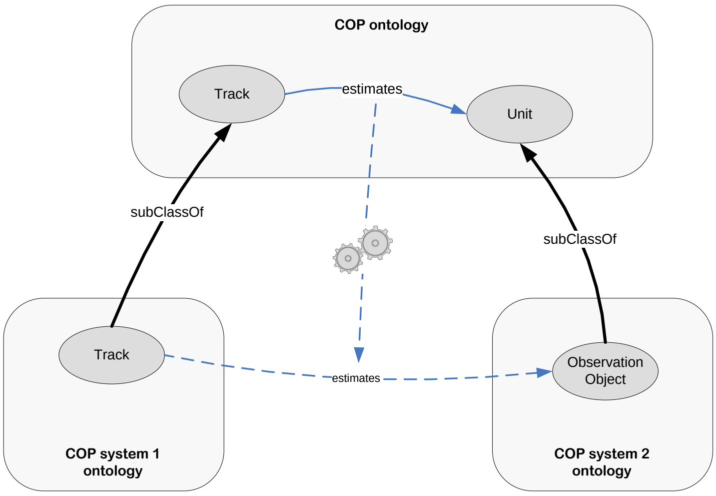

Figure 4: Two COP system ontologies with links to a common COP ontology are matched using reasoning

ontologies O_a (*COP system 1 ontology*) and O_b (*COP system 2 ontology*) are to be matched. Ontology O_a has a concept *Track* (`a:Track`) which through a previous ontology matching process is aligned to be a subclass of the concept *Track* from a common domain ontology (O_c; *COP ontology*) (`c:Track`). This alignment is fetched from an alignment store. Further, a concept from O_b, *ObservationObject* (`b:ObservationObject`) is aligned to be a subclass of the concept *Unit* in O_c (`c:Unit`). Further, in O_c there is specified a relation between the concepts *Track* and *Unit*, indicating that the latter is an estimate of the former:

```
c:Track c:estimates c:Unit.
```

In this case, a reasoner can automatically deduce that the relation `c:estimates` also must exist between the concepts `a:Track` and `b:ObservationObject`:

$$
\begin{array}{c}
(\texttt{a:Track subClassOf c:Track}) \\
\wedge \\
(\texttt{b:ObservationObject subClassOf c:Unit}) \\
\wedge \\
(\texttt{c:Track c:estimates c:Unit}) \\
\Downarrow \\
(\texttt{a:Track c:estimates b:ObservationObject})
\end{array}
$$

We propose that populating the alignment store should be a headquarters staff task, as also illustrated in Figure 3. The reason for this is the quality requirements on the alignment: as mentioned above, the quality of the alignments resulting from deductive ontology matching is dependent on high-quality alignments as input. It is not realistic to expect the users in the theatre in need of information integration to have the required knowledge, nor time, to perform ontology matching of the needed quality.

It is possible to envision that alignments resulting from ontology matching processes performed in theatre could also be an input to the alignment store. However, in order to make sure of the quality of these alignments, we see the need for a review process in the headquarters staff in such a case.

An obvious weakness of this approach is that it cannot guarantee that all possible links between the ontologies in question will be found, as it is totally dependent on the previously stored alignments. This shortcoming will however be alleviated in time, when more and more ontologies have been matched and the resulting alignments have been stored in the alignment store. It is important to keep in mind, however, that the alternative when trying to fetch information from an unanticipated information source often will be no information integration at all. Thus even a few integration links can be of value.

5 Related Work

The literature on ontology matching is vast, with Rahm & Bernstein 2001, Ehrig 2007, and Euzenat & Shvaiko 2007 providing good overviews. Euzenat & Shvaiko 2007 also introduces and gives a good overview of deductive ontology matching.

Euzenat et al. were the first to provide software for storing and sharing alignments with the Alignment API (Euzenat 2004) and the Alignment Server (Euzenat 2008). Their work has to a large degree inspired the ideas presented in this paper.

COMA (++) (Aumueller et al. 2005, Do & Rahm 2002) is an ontology matching system with a focus on reuse of previous alignments. However, it concentrates on the reuse of alignments as a support to a user rather than to use it as a basis for an automated matching system.

Braines et al. 2008 focus on coalition operations and the timeliness issues arising when performing ontology matching. The issue is solved by reusing enhanced alignments: well-defined self-contained ontology fragments created by performing reasoning on the original alignments and their respective ontologies.

Aleksovski et al. 2006 describe an approach very close to our proposal where they seek to create mappings between two (simple) ontologies by first finding correspondences to a common upper ontology, and then deduce the mappings between the ontologies taking the relations in the upper ontology into consideration. The main difference to our approach is that their approach requires an upper ontology. The approach in this paper will make do with any previous alignments involving the ontologies, not just alignments involving an upper ontology.

6 Conclusion and Further Work

The need for flexible solutions to the information integration challenge in NATO Network-Enabled Capability (NNEC) is great in order to fulfil the NNEC goal of improved information sharing among military units. In this paper, deductive ontology matching has been proposed as a support to information integration in time-critical situations, as it has the potential to facilitate the making of automatic ontology matching systems with high-quality output.

The solution proposed in the paper is still at a very early stage, and there remains several questions that need to be answered before it can be held forward as a viable solution to automated ontology matching for information integration in time-critical situations.

The immediate plans for further work in this area include testing the proposed solution in a lab environment using the following existing software components:

- Alignment API (Euzenat 2004)
- Alignment Server (Euzenat 2008)
- Ontologies expressed in the Web Ontology language (OWL) - a World Wide Web Consortium (W3C) recommendation (W3C 2004*a*)
- Collecting information from information sources using W3C's SPARQL Protocol and RDF Query Language (SPARQL) (W3C 2008)
- Data expressed in Resource Description Framework (RDF) (W3C 2004*b*)

In addition to technological try-outs, there also are several scientific challenges connected to this approach that needs to be looked into:

- Is the proposed solution really fast enough for time-critical situations? This question needs to be thoroughly analysed and tested.
- How robust is the proposed solution with respect to the quality of the input alignments? This is a critical question, as the quality of the final alignment is highly dependent on the input alignments when performing deductive ontology matching.
- How can the users supposed to provide the input alignments be supported? What tools do they need?
- How can existing ontology matching algorithms best be combined to give proper support to the users providing the input alignments?

Our further way ahead with this approach is to begin with the technical try-outs. After collecting results regarding the time and robustness issues, we foresee to look further into the issues of user tools and how to combine existing ontology matching algorithms to properly assist ontology matching supervisors.

Acknowledgements

The ideas presented in this paper are based on collaborative work on semantic technologies at the Norwegian Defence Research Establishment, where my colleagues Jonas Halvorsen, Svein Ivar Kristiansen, and Marianne Rustad have made substantial contributions. Also, I would like to thank the reviewers of the paper for their valuable feedback.

References

Aleksovski, Zharko, Michel Klein, Warner ten Kate & Frank Van Harmelen 2006, Matching unstructured vocabularies using a background ontology, *in* 'Proceedings of the 15th International Conference on Knowledge Engineering and Knowledge Management (EKAW)', number 4248 *in* 'Lecture notes in computer science', Springer, pp. 182–197.

Aumueller, David, Hong-Hai Do, Sabine Massmann & Erhard Rahm 2005, Schema and ontology matching with COMA++, *in* 'Proceedings of the 2005 ACM SIGMOD international conference on Management of data', ACM, pp. 906–908.

Braines, Dave, Paul R. Smart, Jie Bao, Alistair Russell, Nigel Shadbolt & James Handler 2008, Using Semantic Web technologies to support information processing and coalition decision making, Technical report, IBM.

Buckman, T. 2005, NATO Network Enabled Capability Feasibility Study Executive Summary. Version 2.0, Technical report, NATO Consultation, Command and Control Agency.

de Bruijn, Jos 2004, Semantic Information Integration Inside and Across Organizational Boundaries, DERI Technical Report, Digital Enterprise Research Institute.

Do, H.H. & E. Rahm 2002, COMA: a system for flexible combination of schema matching approaches, *in* 'Proceedings of the 28th international conference on Very Large Data Bases', VLDB Endowment, pp. 610–621.

Ehrig, Marc 2007, *Ontology Alignment Bridging the Semantic Gap*, Semantic Web and Beyond: Computing for Human Experience, Springer.

Euzenat, Jèrôme 2004, An API for ontology alignment, *in* 'Proc. 3rd conference on international semantic web conference (ISWC)', number 3298 *in* 'Lecture notes in computer science', pp. 698 – 712.

Euzenat, Jèrôme 2008, Alignment API and server (Version 3.2), Technical report, INRIA.

Euzenat, Jèrôme & Pavel Shvaiko 2007, *Ontology Matching*, Springer-Verlag, Heidelberg (DE).

Hansen, Bjørn Jervell 2008, Application of Semantic Technologies in Network Based Defence, *in* 'Proceedings of the 13th International Command and Control Research and Technology Symposium'.

Hitzler, Pascal, Markus Krötzsch & Sebastian Rudolph 2009, *Foundations of Semantic Web Technologies*, CRC Press.

Rahm, Erhard & Philip A. Bernstein 2001, 'A survey of approaches to automatic schema matching', *The VLDB Journal* **10**, 334–350.

Stuckenschmidt, Heiner, Christine Parent & Stefano Spaccapietra, eds 2009, *Modular Ontologies*, Vol. 5445 of *Lecture Notes in Computer Science*, Springer.

W3C 2004*a*, 'OWL Web Ontology Language Overview', http://www.w3.org/TR/owl-features/.

W3C 2004*b*, 'RDF Primer', http://www.w3.org/TR/rdf-primer/.

W3C 2008, 'SPARQL Query Language for RDF', http://www.w3.org/TR/rdf-sparql-query/.