

A Proposal for Optimizing Internetwork Matching of Ontologies

Fabio Santos, Kate Revoredo* and Fernanda Baião**

Department of Applied Informatics
Federal University of the State of Rio de Janeiro (Unirio), Rio de Janeiro, Brazil
{fabio.santos, katerevored, fernanda.baiao}@uniriotec.br

Abstract. A System-of-Systems (SoS) is a set of independent information systems that must communicate with each other towards providing a specific service. Therefore, effectively integrating these systems is demanding. Considering that each system is conceptually described by a unique ontology, the conceptual support for the whole SoS demands the alignment of all ontologies, deriving a network of ontologies. Existing ontology matching techniques may be used for the task; however, due to the recently increasing size of the ontologies and the potential number of ontologies being aligned, current approaches may suffer from scalability and performance issues. In this paper, we introduce an approach to reduce the number of potential correspondences, therefore optimizing the process of creating a network of ontologies. A preliminary experiment was conducted, showing the potential of the proposed approach.

Keywords: network of ontologies · network matching · data integration

1 Introduction

A System-of-Systems (SoS) is defined as a set of independent information systems (IS), providing functionalities derived from the interoperability among them [2]. The development and research on SoS have been gaining increasing attention due to the relevance of several domains such as smart cities, health, emergency response systems, and crisis management systems [6]. Considering that each IS within a SoS is conceptually described by a unique ontology describing its domain, the conceptual support for the whole SoS demands the interoperation of its composing IS, thus requiring the alignment of all the corresponding ontologies. Moreover, a single SoS may embrace several domains, thus requiring by itself a *network of ontologies* as its conceptual support. Therefore, there is an increasing need for aligning networks of ontologies, a problem called *internetwork matching*.

Traditional solutions for ontology matching may be applied for solving the internetwork matching problem, either using a pairwise or a holistic strategy [8]. While the former interactively matches one pair of ontologies from different

* Partially funded by Unirio (PQ-UNIRIO N01/2018)

** Partially funded by CNPq (401505/2014-6)

networks at a time, the latter considers all the network at once. In both cases, all pairs of entities from each ontology that composes the networks are analyzed, which poses a severe restriction in terms of scalability since the required number of comparisons for computing the alignments grows exponentially to the number and size of each ontology. Therefore, there is a need for optimized solutions to this problem [9].

This work proposes an optimized approach for the internetwork matching challenge [1] that tries to reduce the number of pairs to be evaluated during the matching process, thus avoiding unneeded computation while preserving the alignment quality. We evaluated the proposed approach in a preliminary experiment using an OAEI dataset.

This work is organized as follows: Section 2 defines the internetwork matching problem. Our proposed approach is detailed in Section 3. Preliminary evaluation results are in Section 4. Finally, section 5 concludes and points to future work.

2 Problem Definition

A network of ontologies is formally defined as $\Gamma = \langle \Omega, \Lambda \rangle$, where Ω is a finite set of ontologies and $\Lambda(O, O')$ is a set of alignments between pairs of ontologies belonging to Ω [5]. Given a set of two or more networks of ontologies $\Psi = \{\Gamma_1, \Gamma_2, \dots, \Gamma_n\}$, the internetwork matching problem searches for a final network of ontologies Γ_f resulting from the alignments of the networks in Ψ . For instance, Figure 1 depicts two networks of ontologies, each one with 3 ontologies, describing two Systems-of-Systems. The goal is to match these two networks, finding a unique network of ontologies.

One of the approaches for matching networks of ontologies is pairwise. Given a set of networks of ontologies, the pairwise internetwork matching sequentially computes the alignment of each pair of ontologies from each pair of networks from this set. For example, given two networks of ontologies $\Gamma = \langle \Omega, \Lambda \rangle$ and $\Gamma' = \langle \Omega', \Lambda' \rangle$, in which $\Omega = \{O_1, O_2\}$ and $\Omega' = \{O_3\}$, the pairwise internetwork matching is obtained by computing $((O_1 \times O_2) \cup (O_1 \times O_3)) \cup (O_2 \times O_3)$. That is, the pairwise internetwork matching approach computes all matchings between all pairs of ontologies inside each network that is being aligned.

Networks frequently have isomorphisms and trivial alignments that may cause the pairwise approach to find the same alignments more than once, thus requiring an additional step to merge the resulting matches at the end. In the case of isomorphisms, identical correspondences between same entities may be generated (for instance, in Figure 1 a matcher tool may find $A_{1,1'} = \{\langle O_1.a_1, O'_1.a'_1, = \rangle, \langle O_1.b_1, O'_1.b'_1, = \rangle\}$). The case of a trivial alignment occurs when a group of entities, that was previously aligned in a network, appears in another network. For instance, a pairwise matcher that receives the networks Γ and Γ' and has previously computed the intra-network alignments $A_{1,2} = \langle O_1.b_1, O_2.d_2, \sqsubseteq \rangle$ and $A_{1',2'} = \langle O'_1.b'_1, O'_2.d'_2, \sqsubseteq \rangle$ will work unnecessarily to produce $A_{1,2'} = \langle O_1.b_1, O'_2.d'_2, \sqsubseteq \rangle$ and $A_{1',2} = \langle O'_1.b'_1, O_2.d_2, \sqsubseteq \rangle$, which are trivial alignments.

The main weakness of pairwise approach is the number of comparisons needed to compute all alignments and the lack of ability to handle isomorphisms and intra-network alignments.

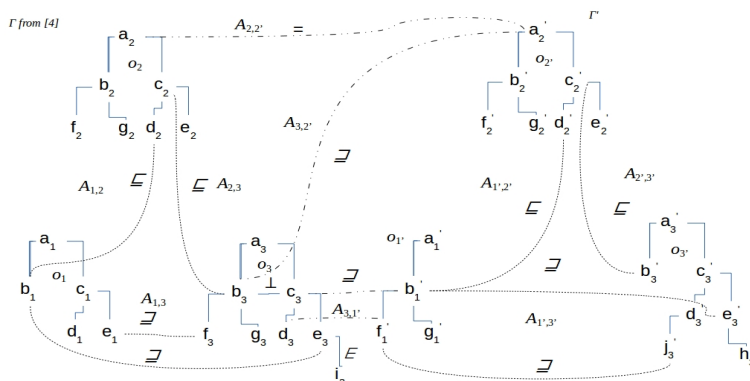


Fig. 1: Internetwork matching of networks Γ and Γ' .

3 Proposed Approach

Given the limitations discussed in Section 2, we propose SubInterNM, a new subsumed approach for the Internetwork matching problem. SubInterNM avoids unnecessary computation by identifying and reusing trivial and subsumed alignments already computed in the networks of ontologies that are being aligned. In such cases, as the intersection among the networks becomes larger, the set of evaluated correspondences during the alignment process tends to get smaller.

For instance, consider the example depicted in Figure 1. The networks Γ and Γ' were aligned using an internetwork matcher. Since O_2 and O_2' are identical, they do not need to be exhaustively compared. Also, both pairs of ontologies O_1 and O_1' and O_3 and O_3' share some subset of entities, thus common parts could be eliminated to compute the final network of ontologies resulted from the internetwork matching problem scenario.

Casanova et al. [3] proposed operations over lightweight ontologies. They define lightweight ontologies as ontologies restricted to DL-Lite core with arbitrary number restrictions. They use a MEG (minimal equivalent graph) approach to create a constraint graph in polynomial time if the graph is acyclic. If the graph is complete, then the problem is NP-Hard. To avoid that, a normalization step is conducted to simplify the graph structure and keep them lightweight.

Our proposed approach SubInterNM uses a combined set of lightweight operations from [3] to verify the existence of isomorphisms. We extrapolate the original idea to use in networks environments, instead of just single ontologies.

4 Preliminary Results

Our proposed approach was evaluated in a preliminary experiment using ontologies from the OAEI conference dataset. We experimented 5 distinct scenarios, and in each of them we specified two networks of ontologies to be aligned. The experiments were defined by increasing the number of ontologies in the network, as well as varying the ontologies and the number of common ontologies between the networks, in order to assess how well our approach would handle the existence of trivial and subsumed alignments:

- 2x2: $\Omega = \{\text{conference, cmt}\}$ and $\Omega' = \{\text{cmt, sigkdd}\}$;
- 3x2: $\Omega = \{\text{conference, cmt, ekaw}\}$ and $\Omega' = \{\text{cmt, sigkdd}\}$;
- 3x3: $\Omega = \{\text{conference, cmt, ekaw}\}$ and $\Omega' = \{\text{cmt, sigkdd, conference}\}$;
- 4x3: $\Omega = \{\text{conference, cmt, dblp, ekaw}\}$ and $\Omega' = \{\text{cmt, sigkdd, conference}\}$;
- 4'x3: $\Omega = \{\text{conference, cmt, edas, ekaw}\}$ and $\Omega' = \{\text{cmt, sigkdd, conference}\}$.

In order to compare our proposed SubInterNM approach against the pairwise internetwork matching approach, we implemented the pairwise approach using the existing matching system ALIN [4] in all experiments. ALIN was selected due to the good results achieved on OEAI 2017, and due to our access to the code [10]. To use ALIN as a *blackbox* (i.e., without having to change its code), for each internetwork matching experiment, we built all the pairs of ontologies to be aligned and interactively invoked ALIN. SubInterNM was implemented

Table 1: Total number of comparisons computed by each approach

Experiment	Pairwise	SubInterNM	% of reduction
2x2	14,138	5,608	60.3
3x2	22,236	10,027	54.9
3x3	38,893	27,039	30.4
4x3	42,319	27,039	36.1
4'x3	57,497	43,420	24.4

using operations defined by Casanova et al. [3], initially considering only the isomorphisms. After, the ALIN matching system [4] was also invoked to compute the alignments between the results from Ω and Ω' . So we computed:

$$\Omega \text{ as } O_1 \cup O_2 \dots \cup O_n - O_1 \cap O'_1 - O_1 \cap O'_2 - \dots - O_n \cap O'_n$$

$$\Omega' \text{ as } O'_1 \cup O'_2 \dots \cup O'_n - O'_1 \cap O_1 - O'_1 \cap O_2 - \dots - O'_n \cap O_n$$

Table 1 shows the number of pairs analyzed (i.e., comparisons) by pairwise and subsumed approach to compute alignments in each experiment. It is possible to verify that SubInterNM reduces the number of comparisons needed to the matching process by at least 24%, therefore representing a successful way to deal with large networks in the internetwork matching problem.

Future work will try to reduce even more the number of comparisons with the detection of intra-network alignments. All the data gathered in the experiment is available at (<https://bit.ly/2M3jIYS>).

5 Conclusion

This work addressed the problem of internetwork ontology matching, a natural evolution of the classical ontology matching problem for highly interconnected scenarios of Systems of Systems, which are of increasing popularity and relevance. We proposed SubInterNM, an approach for internetwork ontology matching that optimizes the required computation of correspondences by identifying and reusing trivial and subsumed alignments. Preliminary evaluation results showed the potential of the approach and opportunities for improvement, in scenarios using lightweight ontologies. The computation of subsumed networks posed an overhead computation since it is a well-known NP-Hard problem [7]. Therefore, further implementations may compute trivial alignments, using it as background knowledge for the matching process [9].

In future work, we expect to improve our implementation, including parallel programming and infrastructure. We also plan to move forward exploring trivial alignments in newer versions of the tool.

References

1. de Abreu Santos, F.M., Revoredo, K., Baião, F.A.: Paving a research roadmap on network of ontologies. In: Proceedings of the 12th International Workshop on Ontology Matching co-located with the 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, October 21, 2017. pp. 221–222 (2017), http://ceur-ws.org/Vol-2032/om2017_poster8.pdf
2. Boehm, B.: A view of 20th and 21st century software engineering. In: Proceedings of the 28th international conference on Software engineering. pp. 12–29. ACM (2006)
3. Casanova, M.A., de Macêdo, J.A., Sacramento, E.R., Pinheiro, Á.M., Vidal, V.M., Breitman, K.K., Furtado, A.L.: Operations over lightweight ontologies. In: OTM Confederated International Conferences” On the Move to Meaningful Internet Systems”. pp. 646–663. Springer (2012)
4. Da Silva, J., Baiao, F.A., Revoredo, K., Euzenat, J.: Semantic interactive ontology matching: synergistic combination of techniques to improve the set of candidate correspondences. In: OM 2017-12th ISWC workshop on ontology matching. pp. 13–24. No commercial editor. (2017)
5. Euzenat, J.: Revision in networks of ontologies. *Artificial intelligence* **228**, 195–216 (2015), <ftp://ftp.inrialpes.fr/pub/exmo/publications/euzenat2015a.pdf>
6. Fitzgerald, J., Foster, S., Ingram, C., Larsen, P.G., Woodcock, J.: Model-based engineering for systems of systems: the compass manifesto. COMPASS Interest Group, Tech. Rep. Manifesto Version **1** (2013)
7. Hsu, H.T.: An algorithm for finding a minimal equivalent graph of a digraph. *Journal of the ACM (JACM)* **22**(1), 11–16 (1975)
8. Megdiche, I., Teste, O., Trojahn, C.: An extensible linear approach for holistic ontology matching. In: International Semantic Web Conference. pp. 393–410. Springer (2016)
9. Shvaiko, P., Euzenat, J.: Ontology matching: state of the art and future challenges. *IEEE Transactions on knowledge and data engineering* **25**(1), 158–176 (2013)
10. da Silva, J., Baiao, F.A., Revoredo, K.: Alin results for oaei 2017. In: OM-2017: Proceedings of the Twelfth International Workshop on Ontology Matching. p. 114 (2017)