

Extending AgreementMakerLight to Perform Holistic Ontology Matching

Marta Contreiras Silva ✉, Daniel Faria, and Catia Pesquita

LASIGE, Faculdade de Ciências da Universidade de Lisboa, Portugal
mcdsilva@fc.ul.pt

Abstract. Creating rich knowledge graphs that allow the representation of data encompassing multiple domains requires the integration of different ontologies. However, the challenge of matching multiple ontologies is not properly addressed by the current pairwise strategy espoused by state-of-the-art ontology alignment systems.

We have extended the ontology alignment system AgreementMakerLight (AML) to address this particular challenge through a scalable cluster-based incremental matching strategy. We make use of AML’s fast and precise matching algorithms to determine the semantic affinity between the ontologies and cluster them, then apply AML’s full ontology matching pipeline incrementally, within each cluster, by matching and then merging ontologies pairwise. The strategy was applied to the integration of 28 biomedical ontologies and achieved a runtime reduction of almost 50%.

This poster expands on the extensions applied to the AML system as the technical contribution that accompanies our In-Use Technology accepted submission ”Matching Multiple Ontologies to Build a Knowledge Graph for Personalized Medicine”.

Keywords: Ontology Matching, Holistic Ontology Matching, Biomedical Ontologies

1 Introduction

Sophisticated Semantic Web applications often depend on a comprehensive and holistic view of their domain. In many domains, there are several ontologies available that provide different scopes or perspectives, and which could be aligned and integrated into a full-fledged semantic network, circumventing the burden of developing a new encompassing ontology. However, this requires a paradigm shift in Ontology Matching from the current pairwise approach of state-of-the-art systems to holistic ontology matching [6].

A naive application of the traditional pairwise alignment strategy would be to align all possible pairs of ontologies and then merge them all to obtain a single alignment. This would require a quadratic number of ontology matching steps, which is not very scalable. As such, the challenge of holistic ontology matching [7, 9] has been addressed in one of two ways: either by partitioning the search space

and then running the pairwise alignments within [4] or matching incrementally in a predefined order [5, 10]. However, these works do not tackle the alignment of ontologies whose domains although related are not exactly the same.

To address the challenge of holistic ontology matching to build a multi-domain graph we have developed a strategy¹ that considers both partitioning and incremental alignment strategies, taking also into account how to support user input. Our strategy aims to strike a balance between quality, coverage and scalability of the alignment process.

2 Holistic OM Strategy

The starting point of our approach is the AML system [3], which has been one of the top performing systems in the Ontology Alignment Evaluation Initiative (OAEI) over the last nine years [8, 1]. However, AML, like other state-of-the-art OM systems is prepared solely to produce alignments between two ontologies, and lacks the functionality of integrating two ontologies through their alignment, which is critical to enable a pairwise matching strategy to be applied incrementally to holistic matching. Our holistic matching strategy, Clustering Incremental

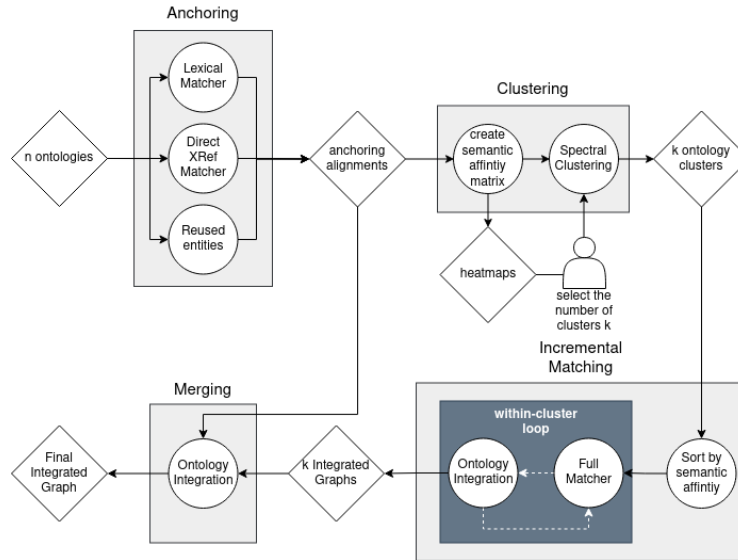


Fig. 1. Overview of the Clustering Incremental Alignment (CIA) strategy.

Alignment (CIA) (Figure 1), is geared to tackle multi-domain matching problems where the overlap between some ontologies might be small or even non-existent.

¹ <https://github.com/liseda-lab/holistic-matching-aml>

For those ontologies, sophisticated matching algorithms would not only be a likely waste of computation time, but could potentially result in lower quality alignments by finding more false positives. However, for highly overlapping ontologies, we do want to use sophisticated matching algorithms, as they are expected to increase the quality of the alignment. Thus, our CIA strategy starts with an anchoring step that aims at defining the level of semantic affinity between all of the ontologies, using fast matching techniques with a high degree of confidence. From these anchoring alignments that are produced, a semantic affinity matrix is created that serves as input for both Spectral Clustering and the development of a heatmap that allows the visualization of more closely related sets of ontologies. This aids the user in defining the number of clusters to select in the Spectral Clustering process. The ontologies are then separated into clusters and further sorted in descending order of semantic affinity. The full AML matching pipeline is then run within each cluster in an incremental fashion, where the first pair of ontologies is matched, then the ontologies are merged through their alignment, then the resulting merged ontology is matched with the third ontology, and so on and so forth. This process results in one integrated graph per cluster, which will be merged together using the initial anchoring alignments to produce a final integrated graph.

3 Application to the Biomedical Domain

The KATY project² aims to develop an AI-empowered personalized medicine system to use in clinical decision support that is based on a Knowledge Graph to support Explainable AI.

To develop this KG we need to integrate 28 biomedical ontologies which are, most often than not, incredibly complex and large. This means that in addition to the challenges in Holistic OM, we are also tackling the challenges in aligning biomedical ontologies [2]. We applied the CIA strategy over this set of 28 ontologies and compared it to the standard AML pairwise strategy, where all possible pairs are matched and then all resulting alignments merged, henceforth referred to as Global Pairwise Alignment (GPA)[11]. Table 1 summarizes the runtimes, number of mappings, and number of tasks from both strategies.

The GPA strategy runs in almost 32 hours, while CIA only requires roughly 16 hours, which corresponds to a decrease in runtime of almost half. Although CIA requires longer loading times (since ontologies are loaded for Anchoring and for Incremental Matching), it spends only 15% of the time in Matching tasks that is required for GPA. The integrated graph produced by CIA is 25% smaller than the one produced by GPA, which can be explained by the lack of semantically redundant mappings that are produced when using basic pairwise matching.

² <http://katy-project.eu/>

Table 1. Alignment results

Strategy	Runtime (hh:mm)			Alignment	
	Load	Match	Total	Mappings	Tasks
GPA	11:47	19:51	31:37	554,547	378
CIA	12:52	03:04	15:56	417,131	378* + 24

GPA: global pairwise alignment. CIA: within cluster incremental alignment.

* from the Anchoring step

4 Conclusions

The CIA strategy decreased the overall runtime required in matching these 28 biomedical ontologies when compared to the state-of-the-art, which makes progress in addressing the issue of scalability. Moreover, the combination of approaches in CIA is expected to ensure high precision across less closely related domains (different clusters) by employing high precision lexical matching, while improving recall for closely related ontologies (same cluster) by employing more computationally expensive algorithms. This expectation is supported by the evaluation of the matching algorithms of AML in the OAEI campaign where the full AML strategy achieves median performances of 90.6% in precision and 76.31% in recall in the anatomy and large biomed tracks, whereas the lexical matching achieves 98.9% precision and 40.6% recall [1, 2].

The next step in this work is to perform an expert-based validation study over a set of selected mappings to ensure that the produced alignment is correct and coherent. The use of this type of validation is vital since the available reference alignments are incapable of supporting our alignments tasks.

Acknowledgments This work was supported by FCT through the LASIGE Research Unit (UIDB/00408/2020 and UIDP/00408/2020). It was also partially supported by the KATY project funded the European Union’s Horizon 2020 research and innovation program (GA 101017453).

References

1. Faria, D., Lima, B., Silva, M.C., Couto, F.M., Pesquita, C.: Aml and amlc results for oaei 2021 p. 6
2. Faria, D., Pesquita, C., Mott, I., Martins, C., Couto, F.M., Cruz, I.F.: Tackling the challenges of matching biomedical ontologies. *Journal of biomedical semantics* **9**(1), 1–19 (2018)
3. Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I.F., Couto, F.M.: The agreementmakerlight ontology matching system. In: *OTM Confederated International Conferences” On the Move to Meaningful Internet Systems”*. pp. 527–541. Springer (2013)

4. Gruetze, T., Böhm, C., Naumann, F.: Holistic and scalable ontology alignment for linked open data. *LDOW* **937**, 1–10 (2012)
5. Hertling, S., Paulheim, H.: Order matters: Matching multiple knowledge graphs. arXiv preprint arXiv:2111.02239 (2021)
6. Megdiche, I., Teste, O., Trojahn, C.: An extensible linear approach for holistic ontology matching. In: *International Semantic Web Conference*. pp. 393–410. Springer (2016)
7. Otero-Cerdeira, L., Rodríguez-Martínez, F.J., Gómez-Rodríguez, A.: Ontology matching: A literature review. *Expert Systems with Applications* **42**(2), 949–971 (2015)
8. Pour, M.A.N., Algergawy, A., Amardeilh, F., Amini, R., Fallatah, O., Faria, D., Fundulaki, I., Harrow, I., Hertling, S., Hitzler, P., Huschka, M., Ibanescu, L., Jimenez-Ruiz, E., Karam, N., Laadhar, A., Lambrix, P., Li, H., Li, Y., Michel, F., Nasr, E., Paulheim, H., Pesquita, C., Portisch, J., Roussey, C., Saveta, T., Shvaiko, P., Splendiani, A., Trojahn, C., Vataščinová, J., Yaman, B., Zamazal, O.: Results of the ontology alignment evaluation initiative 2021
9. Rahm, E.: Towards large-scale schema and ontology matching. In: *Schema matching and mapping*, pp. 3–27. Springer (2011)
10. Saleem, K., Bellahsene, Z., Hunt, E.: Porsche: Performance oriented schema mediation. *Information Systems* **33**(7-8), 637–657 (2008)
11. Silva, M.C., Faria, D., Pesquita, C.: Matching multiple ontologies to build a knowledge graph for personalized medicine. In: *International Semantic Web Conference 2022*