Towards Tailoring Ontology Embeddings for Ontology Matching Tasks

Sevinj Teymurova[1,*], Ernesto Jiménez-Ruiz[1,2], Tillman Weyde[1] and Jiaoyan Chen[3]

[1]*City St George's, University of London, UK*
[2]*University of Oslo, Norway*
[3]*The University of Manchester, UK*

Abstract
Ontology alignment becomes crucial for achieving semantic interoperability as the multiple ontologies representing the same domain are increasing. This paper introduces OWL2Vec4OA, an enhancement of the OWL2Vec* ontology embedding system. Although OWL2Vec* is a robust method for ontology embedding, it currently lacks specialization for ontology alignment tasks. OWL2Vec4OA addresses this limitation by incorporating confidence values from seed mappings to bias its random walk approach.

Keywords
ontology alignment, walking strategy, ontology embeddings, knowledge graph embeddings

1. Introduction and System Overview

Ontology alignment plays a key role in achieving semantic interoperability and enhancing the use of knowledge graphs. The Ontology Alignment Evaluation Initiative (OAEI) [1, 2] has driven progress in this field for two decades with the systematic evaluation of ontology alignment systems, that is, systems aiming at identifying semantic relationships between different ontologies [3]. Recent advancements incorporate knowledge graph embeddings, machine learning and large language models, categorized into: *(i)* direct embedding comparison (e.g., ERSOM [4], DeepAlignment [5]); *(ii)* supervised mapping classifiers (e.g., VeeAlign [6], MEDTO [7], LogMap-ML [8], SORBET [9]); and *(iii)* language model-based approaches (e.g., BERTMap [10], BERTSubs [11], OLaLa [12])

This paper summarizes OWL2Vec4OA,[1] an extension of the ontology embedding system OWL2Vec* [14] tailored to the ontology alignment task. The main steps of our OWL2Vec4OA are depicted in Figure 1 and summarised as follows. *(i) Ontology Projection:* We employ the same projection rules as OWL2Vec* to transform ontology axioms into RDF triples, creating a directed labeled graph for each input ontology. *(ii) Ontology Alignment:* We utilize traditional ontology matching systems, LogMap [15] and AML [16], to produce seed mappings. These mappings bridge the ontology graphs, enabling random walks across entities from different ontologies. *(iii) Graph Merger and Edge Weighting:* Unlike OWL2Vec*, OWL2Vec4OA constructs a single (weighted) graph from the ontology projections and seed mappings. Edges derived from ontology axioms are assigned a weight of 1.0, while edges from mappings are weighted according to their confidence values. *(iv) Biased Random Walks:* Inspired by Cochez et al. [17], we implement a biased walk strategy that considers the weight of the edges. *(v) Document Generation:* OWL2Vec4OA generates three types of documents by replacing in the created entity sequences some, none or all URI mentions by their lexical representation. *(vi) Word2vec Application:* We apply the Word2vec model [18] to compute embeddings for both URIs and words.

Our predecessor OWL2Vec*, although has a setting where multiple ontologies can be given as input, focuses on the embedding of single ontologies. OWL2Vec4OA, however, allows for a tight connection

OM-2024: The 19th International Workshop on Ontology Matching collocated with the 23rd International Semantic Web Conference (ISWC 2024), November 11th, Baltimore, USA
sevinj.teymurova@city.ac.uk (S. Teymurova); Ernesto.Jimenez-Ruiz@city.ac.uk (E. Jiménez-Ruiz); T.E.Weyde@city.ac.uk (T. Weyde); jiaoyan.chen@manchester.ac.uk (J. Chen)
https://github.com/Sevinjt/OWL2Vec4OA (S. Teymurova)

[1]A longer version of this paper has been submitted to a conference [13].

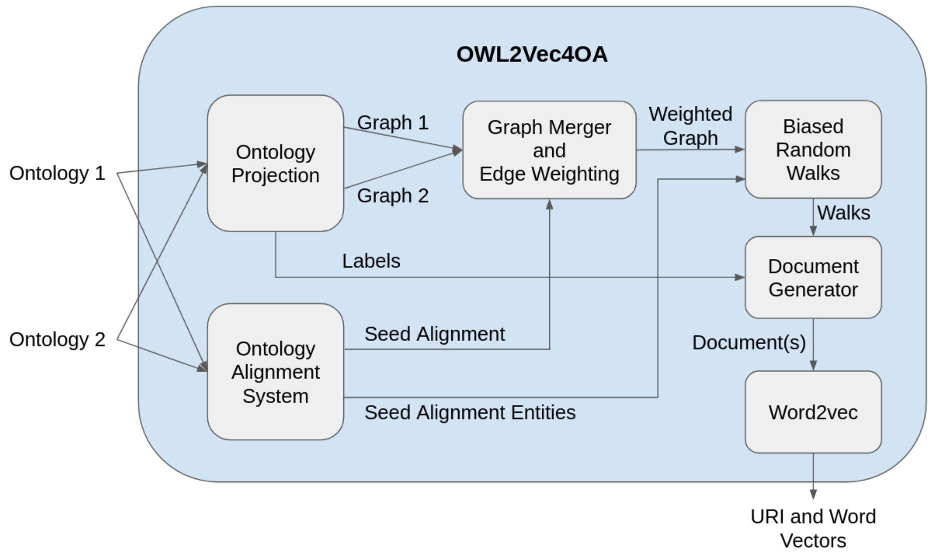

Figure 1: General architecture of OWL2Vec4OA

of the input ontologies given a set of seed mappings, which enables the creation of sequences across entities from different ontologies. The implementation of OWL2Vec4OA is available on our GitHub repository: https://github.com/Sevinjt/OWL2Vec4OA

2. Results, Discussion and Future Work

Our study evaluated the performance of OWL2Vec4OA across multiple biomedical ontology alignment tasks from the local matching setting of the OAEI's Bio-ML track [19]. The local matching tasks consist in ranking a correct mapping among a pool of incorrect mappings.

In this work, mappings are scored and ranked according to the cosine similarity of the computed URI embeddings for the entities in the mapping. We applied OWL2Vec4OA and OWL2Vec* to compute the embeddings, fixing the Word2Vec hyper parameters — the number of epochs and embedding dimension to 70 and 100, respectively. OWL2Vec4OA demonstrated significant improvements over our predecessor OWL2Vec* for all the tested ontology pairs, indicating that the OWL2Vec4OA embeddings are better suited for ontology alignment tasks. For OMIM-ORDO, OWL2Vec4OA showed substantial improvement at walk depth 2, with Mean Reciprocal Rank (MRR) increasing from 0.074 to 0.586, and Hits@1 improving from 0.018 to 0.533. In NCIT-DOID, OWL2Vec4OA achieved its best performance at walk depth 4, with MRR rising from 0.105 to 0.609, and Hits@1 from 0.035 to 0.442. SNOMED-NCIT-N exhibited the most dramatic improvement. At walk depth 4, MRR increased from 0.055 to 0.805, and Hits@1 from 0.011 to 0.747. For SNOMED-NCIT-P, significant improvements were observed at walk depth 2, with MRR increasing from 0.079 to 0.436, and Hits@1 from 0.018 to 0.342. Walk length significantly influenced performance across different ontology pairs. Generally, shorter walk lengths (2 or 3) performed better for some pairs like OMIM-ORDO and SNOMED-NCIT-P, while others such as NCIT-DOID and SNOMED-NCIT-N benefited from longer walk lengths. Computation time varied based on ontology pair and walk depth, with longer depth consistently requiring more time than shorter walk depth.

We plan to extend our work as follows: *(i)* train a machine learning model with OWL2Vec4OA embeddings, similar to approaches like LogMap-ML and Hao et al. [20]; *(ii)* perform additional experiments to better understand the impact of the walk depth with different strategies to create entity sequences (*i.e.*, focusing on concepts and/or avoiding OWL constructs); and *(iii)* create an end-to-end ontology alignment system to participate in the OAEI campaign.

Table 1
Results of OWL2Vec4OA and OWL2Vec* over four Bio-ML tasks, with different walk depths (wd).

Task	System	wd	MRR	Hits@1	Hits@5	Hits@10	Hits@20	Hits@30
OMIM-ORDO	OWL2Vec*	2	0.074	0.018	0.091	0.178	0.332	0.393
		3	0.073	0.018	0.090	0.170	0.318	0.381
		4	0.071	0.019	0.078	0.320	0.321	0.387
	OWL2Vec4OA	2	**0.586**	**0.533**	**0.637**	**0.657**	**0.672**	**0.693**
		3	0.402	0.306	0.512	0.587	0.650	0.685
		4	0.215	0.132	0.281	0.359	0.446	0.532
NCIT-DOID	OWL2Vec*	2	0.218	0.110	0.306	0.448	0.631	0.746
		3	0.175	0.074	0.251	0.377	0.561	0.690
		4	0.105	0.035	0.121	0.225	0.409	0.541
	OWL2Vec4OA	2	0.195	0.064	0.310	0.508	0.709	0.812
		3	0.358	0.181	0.573	0.741	0.872	0.924
		4	**0.609**	**0.442**	**0.840**	**0.928**	**0.970**	**0.984**
SNOMED-NCIT-N	OWL2Vec*	2	0.063	0.014	0.075	0.134	0.231	0.309
		3	0.068	0.017	0.079	0.142	0.238	0.308
		4	0.055	0.011	0.052	0.114	0.218	0.305
	OWL2Vec4OA	2	0.648	0.543	0.767	0.831	0.888	0.904
		3	0.605	0.484	0.746	0.813	0.872	0.899
		4	**0.805**	**0.747**	**0.872**	**0.888**	**0.902**	**0.910**
SNOMED-NCIT-P	OWL2Vec*	2	0.079	0.018	0.094	0.184	0.302	0.675
		3	0.078	0.018	0.092	0.181	0.292	0.667
		4	0.055	0.011	0.052	0.114	0.218	0.305
	OWL2Vec4OA	2	**0.436**	**0.342**	**0.534**	**0.583**	**0.609**	**0.967**
		3	0.311	0.190	0.435	0.502	0.558	0.933
		4	0.291	0.204	0.355	0.434	0.521	0.944

Acknowledgments

This research is funded by the Ministry of Education and Science of Azerbaijan Republic with support from City St George's, University of London. This work has also been partially supported by the Academy of Medical Sciences Network Grant (Neurosymbolic AI for Medicine, NGR1\1857) and the project "XAI4SOC: Explainable Artificial Intelligence for Healthy Aging and Social Wellbeing" funded by the Agencia Estatal de Investigación (AEI), the Spanish Ministry of Science, Innovation and Universities and the European Social Funds (PID2021-123152OB-C22).

References

[1] M. A. N. Pour, et al., Results of the ontology alignment evaluation initiative 2022, in: Proceedings of the 17th International Workshop on Ontology Matching (OM), volume 3324 of *CEUR Workshop Proceedings*, 2022, pp. 84–128. URL: https://ceur-ws.org/Vol-3324/oaei22_paper0.pdf.

[2] M. A. N. Pour, et al., Results of the Ontology Alignment Evaluation Initiative 2023, in: Proceedings of the 18th International Workshop on Ontology Matching (OM), volume 3591 of *CEUR Workshop Proceedings*, 2023, pp. 97–139. URL: https://ceur-ws.org/Vol-3591/oaei23_paper0.pdf.

[3] L. Otero-Cerdeira, F. J. Rodríguez-Martínez, A. Gómez-Rodríguez, Ontology matching: A literature review, Expert Systems with Applications 42 (2015) 949–971.

[4] C. Xiang, T. Jiang, B. Chang, Z. Sui, ERSOM: A structural ontology matching approach using

automatically learned entity representation, in: Proceedings of the 2015 conference on empirical methods in natural language processing, 2015, pp. 2419–2429.
[5] P. Kolyvakis, A. Kalousis, D. Kiritsis, DeepAlignment: Unsupervised ontology matching with refined word vectors, in: Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1-6 June 2018, 2018.
[6] V. Iyer, A. Agarwal, H. Kumar, VeeAlign: Multifaceted Context Representation Using Dual Attention for Ontology Alignment, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, 2021, pp. 10780–10792. doi:10.18653/V1/2021.EMNLP-MAIN.842.
[7] J. Hao, C. Lei, V. Efthymiou, A. Quamar, F. Özcan, Y. Sun, W. Wang, Medto: Medical data to ontology matching using hybrid graph neural networks, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021, pp. 2946–2954.
[8] J. Chen, E. Jiménez-Ruiz, I. Horrocks, D. Antonyrajah, A. Hadian, J. Lee, Augmenting ontology alignment by semantic embedding and distant supervision, in: The Semantic Web: ESWC, Springer, 2021, pp. 392–408.
[9] F. Gosselin, A. Zouaq, SORBET: A Siamese Network for Ontology Embeddings Using a Distance-Based Regression Loss and BERT, in: International Semantic Web Conference, Springer, 2023, pp. 561–578.
[10] Y. He, J. Chen, D. Antonyrajah, I. Horrocks, BERTMap: A BERT-Based Ontology Alignment System, in: Thirty-Sixth AAAI Conference on Artificial Intelligence, 2022.
[11] J. Chen, Y. He, Y. Geng, E. Jiménez-Ruiz, H. Dong, I. Horrocks, Contextual semantic embeddings for ontology subsumption prediction, World Wide Web 26 (2023) 2569–2591.
[12] S. Hertling, H. Paulheim, OLaLa: Ontology Matching with Large Language Models, in: K. B. Venable, D. Garijo, B. Jalaian (Eds.), Proceedings of the 12th Knowledge Capture Conference (K-CAP), ACM, 2023, pp. 131–139. URL: https://doi.org/10.1145/3587259.3627571. doi:10.1145/3587259.3627571.
[13] S. Teymurova, E. Jiménez-Ruiz, T. Weyde, J. Chen, OWL2Vec4OA: Tailoring Knowledge Graph Embeddings for Ontology Alignment, in: Submitted to a Conference, 2024. Paper available here: http://arxiv.org/abs/2408.06310.
[14] J. Chen, P. Hu, E. Jimenez-Ruiz, O. M. Holter, D. Antonyrajah, I. Horrocks, OWL2vec*: Embedding of OWL ontologies, Machine Learning 110 (2021) 1813–1845.
[15] E. Jimenez-Ruiz, B. Cuenca Grau, LogMap: Logic- Based and Scalable Ontology Matching, The Semantic Web – ISWC (2011) vol 7031.
[16] D. Faria, C. Pesquita, E. Santos, M. Palmonari, I. F. Cruz, F. M. Couto, The agreementmakerlight ontology matching system, in: On the Move to Meaningful Internet Systems, volume 8185 of *Lecture Notes in Computer Science*, Springer, 2013, pp. 527–541. URL: https://doi.org/10.1007/978-3-642-41030-7_38. doi:10.1007/978-3-642-41030-7_38.
[17] M. Cochez, P. Ristoski, S. P. Ponzetto, H. Paulheim, Biased graph walks for RDF graph embeddings, in: R. Akerkar, A. Cuzzocrea, J. Cao, M. Hacid (Eds.), Proceedings of the 7th International Conference on Web Intelligence, Mining and Semantics, ACM, 2017, pp. 21:1–21:12. URL: https://doi.org/10.1145/3102254.3102279. doi:10.1145/3102254.3102279.
[18] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, Advances in neural information processing systems 26 (2013).
[19] Y. He, J. Chen, H. Dong, E. Jiménez-Ruiz, A. Hadian, I. Horrocks, Machine Learning-Friendly Biomedical Datasets for Equivalence and Subsumption Ontology Matching, in: 21st International Semantic Web Conference, volume 13489 of *Lecture Notes in Computer Science*, Springer, 2022, pp. 575–591. URL: https://doi.org/10.1007/978-3-031-19433-7_33. doi:10.1007/978-3-031-19433-7_33.
[20] Z. Hao, W. Mayer, J. Xia, G. Li, L. Qin, Z. Feng, Ontology alignment with semantic and structural embeddings, J. Web Semant. 78 (2023) 100798. URL: https://doi.org/10.1016/j.websem.2023.100798. doi:10.1016/J.WEBSEM.2023.100798.