Agent-OM Results for OAEI 2025

Zhangcheng Qiang[1,*], Weiqing Wang[2] and Kerry Taylor[1]

[1]Australian National University, School of Computing, 108 North Road, Acton, ACT 2601, Canberra, Australia
[2]Monash University, Faculty of Information Technology, 25 Exhibition Walk, Clayton, VIC 3800, Melbourne, Australia

Abstract

We present the results obtained in the Ontology Alignment Evaluation Initiative (OAEI) 2025 campaign using our ontology matching (OM) system Agent-OM. This is our first participation in the OAEI campaign, featuring two variants with different large language models (LLMs): The *production* version uses commercial LLMs for optimal performance, while the *lite* version uses open-source LLMs for cost-effectiveness. Experimental results in eight OAEI tracks demonstrate the generative power of Agent-OM in handling OM tasks from diverse domains, languages, and vocabularies. We also outline future directions to improve our system.

Keywords

ontology matching, OAEI campaign

1. Presentation of the system

Agent-OM is one of the leading systems for ontology matching (OM). Leveraging the power of large language models (LLMs) and the advantages of LLM agents, the system achieves performance comparable to state-of-the-art systems in simple OM tasks and surpasses them in complex and few-shot OM tasks [1].

1.1. New features of Agent-OM 2025 edition

- Agent-OM has been extended for better functionality and usability.
  - We develop Agent-OM-Lite, a free and lite version for cost-effectiveness in large-scale OM.
  - We extend Agent-OM for ontology versioning [2] and LLM hallucination evaluation [3].
  - We extend Agent-OM to support the latest DeepSeek models [4] and OpenAI open-source models [5].
  - We extend the precision-recall-F1 chart used in Agent-OM to an interactive user tool [6].
- Agent-OM has been re-engineered to support OM tasks lacking a reference alignment. We now allow users to upload only the source and target ontologies. If the system detects that no reference alignment exists, it will create a dummy one to enable the pipeline. The alignment generated by Agent-OM itself will subsequently replace the dummy reference file.
- Agent-OM has been upgraded to interpret OWL restrictions. The key challenge is that these restrictions contain nested logical expressions with blank nodes. The new function is attached to the semantic retriever and now the retriever can capture both hierarchical and logical relationships. Packaged in an LLM-based tool, it can be connected for complex tasks and unplugged for tasks where it is not required.
- Agent-OM has been optimised for reproducibility. We employ several new hyperparameters to instruct LLMs to return stable output over multiple runs. We use $top\_k = 1$ and $top\_p = 1.0$ to ensure that LLMs always choose the top-one token and do not filter the output. Note that these parameters may not be modifiable in commercial LLMs. For example, the top_k value is currently not available in OpenAI models. We use $repeat\_penalty = 1.0$ (or similar parameters, $presence\_penalty = 0.0$ and $frequency\_penalty = 0.0$) to assign no penalty for repeated output from LLMs; in other words, to encourage LLMs to produce repetitive and stable outputs.

OM 2025: The 20th International Workshop on Ontology Matching collocated with the 24th International Semantic Web Conference (ISWC 2025), November 2nd, 2025, Nara, Japan

*Corresponding author.
qzc438@gmail.com (Z. Qiang); teresa.wang@monash.edu (W. Wang); kerry.taylor@anu.edu.au (K. Taylor)
0000-0001-5977-6506 (Z. Qiang); 0000-0002-9578-819X (W. Wang); 0000-0003-2447-1088 (K. Taylor)

1.2. Agent-OM in the OAEI 2025 campaign

There are two Agent-OM variants that participate in the OAEI 2025 campaign.

• **Agent-OM** is the production version of Agent-OM. The backend uses commercial LLMs and the corresponding embedding models. The production version achieves optimal performance, but requires extensive access to commercial APIs. The results show slight differences across different runs due to limited support for reproducibly fixing the model's hyperparameters.

• **Agent-OM-Lite** is the lite version of Agent-OM. The backend uses open-source LLMs for both language processing and text embedding. Although the performance of the lightweight version is usually poorer than that of the production version, it offers an alternative solution for cost-constrained or security-constrained scenarios. The result is more stable across different runs.

1.2.1. System settings

Figure 1 provides the LLM variations available for OAEI 2025. For commercial API-accessed LLMs used in Agent-OM, gpt-4o [7] with the timestamp tag 2024-05-13 has the optimal performance and stability. However, its API cost can be expensive for large-scale OM tasks. Alternatives include the late-breaking version without the timestamp tag and the mini version gpt-4o-mini [8]. Note that the late-breaking version may produce less stable results, while the mini version can lower the matching performance. For open-source llama-3 [9] models used in Agent-OM-Lite, the large-size model llama-3-70b can perform better than the small-size model llama-3-8b, but the execution time may be longer.

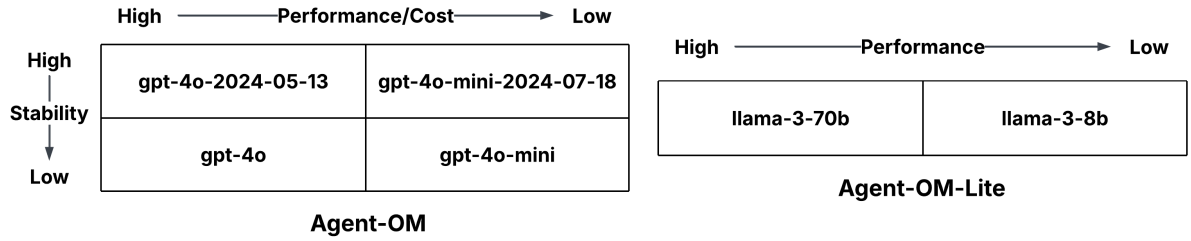

Figure 1: LLM variations available for OAEI 2025.

Table 1 shows the hyperparameter settings for OAEI 2025. We use gpt-4o(-mini) with the text embedding model ada-002 [10] for the production version Agent-OM and llama-3 for Agent-OM-Lite. The global settings of similarity_threshold = 0.90 and top@k = 3 may not be optimal for each track. We recommend trying different settings to find a customised setting for each task. The new LLM hyperparameters may cause additional execution time for LLMs. If the task does not have a reproducibility requirement, we suggest setting the temperature to 0.0 and ignoring other hyperparameters. There is only a slight difference between multiple runs with this setting. The system hyperparameter $top@k$ is used to restrict the top k matching candidates chosen by Agent-OM and its lite version, while the LLM hyperparameter top_k is used to restrict the top k tokens selected by the LLM used in Agent-OM and its lite version. The top_p value is not functional when $top\_k = 1$.

Table 1
Hyperparameter settings for OAEI 2025.

System Hyperparameters		
$similarity\_threshold = 0.90, top@k = 3, seed = 42$		
System Variant	**LLM**	**LLM Hyperparameters**
Agent-OM	gpt-4o(-mini)	$temperature = 0.0, top\_p = 1.0, presence\_penalty = 0.0, frequency\_penalty = 0.0$
	ada-002	$embedding\_vector\_length = 1536$
Agent-OM-Lite	llama-3	$temperature = 0.0, top\_k = 1, repeat\_penalty = 1.0, embedding\_vector\_length = 4096$

Table 2 shows the ontology structural terms interpreted for OAEI 2025. We observe that OAEI tracks mainly use OWL (http://www.w3.org/2002/07/owl#)/RDFS (http://www.w3.org/2000/01/rdf-schema#) and SKOS (http://www.w3.org/2004/02/skos/core#) for their vocabularies. The syntactic retriever processes class and property names, while the lexical retriever processes descriptive terms, and the semantic retriever processes hierarchical and logical terms.

Table 2
Ontology structural terms interpreted for OAEI 2025.

Retriever	OWL/RDFS	SKOS
Syntactic	owl:Class, owl:ObjectProperty, owl:DatatypeProperty rdfs:label,...	skos:Concept skos:prefLabel, skos:altLabel,...
Lexical	rdfs:comment,...	skos:definition, skos:note,...
Semantic	rdfs:subClassOf, rdfs:subPropertyOf,... owl:equivalentClass, owl:equivalentProperty,... rdfs:domain, rdfs:range,... owl:Restriction, owl:onClass, owl:onProperty,...	skos:broader, skos:narrower, skos:related,... skos:exactMatch, skos:closeMatch,... skos:hasTopConcept, skos:topConceptOf,...

1.2.2. Performance reporting

For the confidence of each mapping, Agent-OM provides an approximate range (e.g. $confidence \geq 0.90$), but not the exact value (e.g. $confidence = 0.97$). This is because Agent-OM applies reciprocal rank fusion (RRF) [11] on top of the matching results, and the ranking results do not have a statistical link to confidence. We use RRF to overcome two limitations of the traditional approach by computing a mean of syntactic, lexical, and semantic matching results (as illustrated in Figure 2):

(1) The traditional approach cannot determine the best match between very similar entities. For example, entities may have the same mean value despite having different results in syntactic, lexical, and semantic matching (coloured blue in the figure). By computing and accumulating their rankings, the RRF approach is able to distinguish the best match from other close matches.
(2) The traditional approach is very sensitive to insufficient input data causing semantic matching to fail. For example, an entity with missing results in semantic matching will obtain a very low mean value (coloured red in the figure). In such cases, the RRF approach is able to minimise the impact of missing values so that the entity with missing values becomes comparable with other entities.

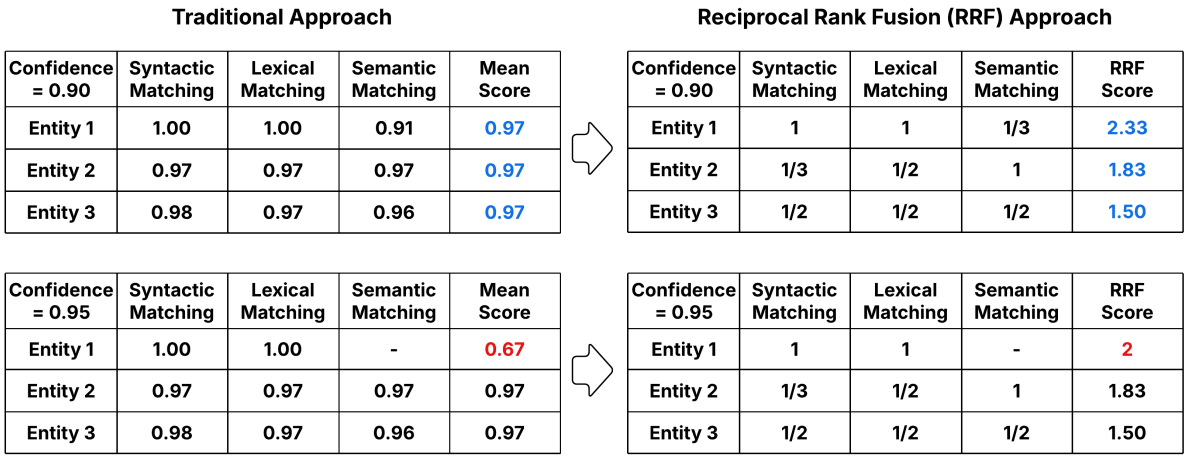

Figure 2: Traditional Approach vs Reciprocal Rank Fusion (RRF) Approach.

Agent-OM is expected to have a longer execution time than traditional OM systems. Agent-OM is built on LLM agents, which are inherently characterised by latency behaviours. Its powerful capability in reasoning is achieved by accumulating historical context and enabling a comprehensive tool-augmented extension. This results in a lengthy context fed into LLMs, as well as increased resource usage in tool calling and memory access [12]. Additionally, for accessing commercial API-accessed models (e.g. gpt-4o and gpt-4o-mini), the execution time is under the control of the API provider and not of the matcher. Guardrails are typically applied to restrict the number of requests per minute (RPM) and tokens per minute (TPM). For example, the OpenAI rate limits given in [13]. For accessing open-source models (e.g. llama-3), the execution time depends on the settings of the local machine. On machines equipped with graphics processing units (GPUs), the processing is significantly faster than on machines with only central processing units (CPUs). Agent-OM has multiple CRUD (create, read, update, and delete) functions on its database. The time used in querying and searching is driven by the choice of database implementation.

1.3. Link to the system and parameters file

Agent-OM is open-source and released under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The source code, data, and/or other artifacts have been made available at https://github.com/qzc438/ontology-llm.

1.4. Link to the system alignments

The system alignments are stored in the folder named OAEI_2025 at https://github.com/qzc438/ontology-llm/tree/master/campaign/. For large datasets, the complete results are stored in the folder named OAEI_2025 at https://github.com/qzc438/ontology-llm-large-datasets/tree/master/campaign/. Under each track folder, predict.csv/predict.xml corresponds to our system alignment, whereas true.csv corresponds to the reference alignment. The confidence for each mapping in predict.csv/predict.xml is greater than or equal to 0.90, produced by the setting of $similarity\_threshold = 0.90$. Note that we do not include the element <measure> in our alignment file, while the evaluation conducted in the Matching EvaLuation Toolkit (MELT) [14] will add <measure rdf:datatype="xsd:float">1.0</measure> in this case. The result.csv file reports the measures of precision, recall, and F1 score. Some rows present intermediate partial results, but rows ending with "llm_with_agent" in the "Alignment"column present the final matching results. The execution time is not reported due to variations in the API provider (for commercial API-accessed LLMs) and computational power (for open-source LLMs). It follows a linear growth with the number of entities if no additional optimisations are applied, such as those in Section 3.2. The cost.csv file reports the API charge for API-accessed LLMs. Open-source LLMs are used free-of-charge.

2. Results

Table 3 shows Agent-OM participation in the OAEI 2025 TBox/schema matching tracks. Agent-OM focuses on one-to-one equivalence mapping in the TBox/schema matching tasks. The current system has limited support for instance matching and link discovery. We do not participate in the TBox/schema matching tracks that contain interactive matching and complex matching types or relations. We refer readers to the official website for the results of Agent-OM in the OAEI 2025 campaign: https://oaei.ontologymatching.org/2025/results/. Note that the results are produced from a single trial by the authors, and slight differences may occur across multiple runs due to the non-determinism of LLMs. The system variant and chosen LLM are determined by balancing performance and cost efficiency. For small tasks, we use our premium Agent-OM working with the premium gpt-4o-2024-05-13 model, which gives our best results at a high cost. For medium-sized tasks, we use Agent-OM with the inexpensive gpt-4o-mini-2024-07-18 model. For large-scale tasks, we use Agent-OM-Lite with the free-of-charge open-source llama-3-8b model running entirely on our local machine.

Table 3
Agent-OM participation in the OAEI 2025 TBox/schema matching tracks.

Track Name	Track Domain	System Variant	Chosen LLM	Evaluation
anatomy	Human and Mouse Anatomy	Agent-OM	gpt-4o-2024-05-13	complete
conference	Research Conference	Agent-OM	gpt-4o-2024-05-13	complete
multifarm	Conference Extension on Multilingualism	Agent-OM	gpt-4o-mini-2024-07-18	-
bio-ml	Biomedical	Agent-OM-Lite	llama-3-8b	partial
biodiv	Biodiversity and Ecology	Agent-OM-Lite	llama-3-8b	-
dh	Digital Humanities	Agent-OM	gpt-4o-2024-05-13	complete
arch-multiling	Archaeology Multiling	Agent-OM	gpt-4o-2024-05-13	complete
ce	Circular Economy	Agent-OM	gpt-4o-2024-05-13	complete

Tracks labelled "-" were unable to publish Agent-OM results due to platform issues or communication failures.

Given the rank of some matcher $rank_i$ on track i, and the number of participants on that track, our goal is to normalise all reciprocal ranks to a scale of $[0, 1]$, with 1 corresponding to the highest rank and 0 to the lowest. Therefore, the normalised reciprocal rank (RR^*) and the overall mean reciprocal rank (MRR^*) are defined as:

$$RR_i^* = 1 - \frac{(\text{rank}_i - 1)}{(\text{number of participants}_i - 1)} \qquad MRR^* = \frac{1}{N}\sum_{i=1}^{N} RR_i^* \tag{1}$$

Figure 3 shows Agent-OM's normalised RR^* per track and overall MRR^* on the tracks shown as "complete" in Table 3. We can see no pattern in Agent-OM's precision vs recall performance ranking running across the tracks, although this may reflect track-wise precision-recall variability in other matchers. The conference and dh tracks have a notable gap between rankings in precision and recall, suggesting room for improvement. Regarding MRR^*, we observe that Agent-OM's precision and F1 score are very similar, suggesting that Agent-OM could be more competitive by prioritising improvements of recall over precision.

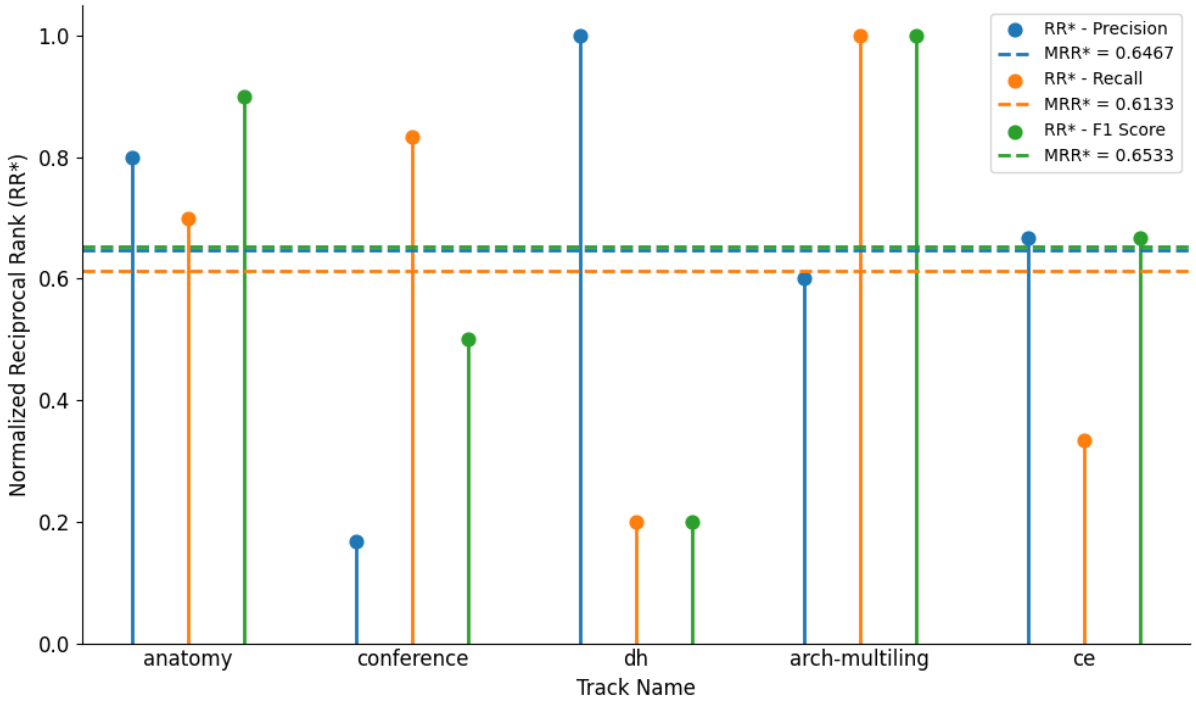

Figure 3: Agent-OM's normalised reciprocal rank (RR^*) per track and overall mean reciprocal rank (MRR^*).

3. General comments and conclusions

3.1. Comments on the results

(1) We apply Agent-OM to three previously evaluated tracks (anatomy, conference, and mse) in [1] and two new tracks (dh and ce). While the mse track is not participating in the 2025 campaign, we provide our results for this track on GitHub for reference (see Section 1.3). The results indicate that Agent-OM is resilient in performing tasks from diverse domains with varying levels of complexity and ontology structural terms used. Although performance by traditional systems on simple tasks remains comparable, we believe that Agent-OM is paving the way for shifting to general-purpose domain-independent OM systems.
(2) We apply Agent-OM to two multilingual tracks (multifarm and arch-multiling). We find that matching ontologies expressed in the same language is more successful than matching different languages, although having English as one of the pair of different languages is clearly advantageous. Further, matches between ontologies using languages from the same language family (e.g. English matched with European languages) are better than those between different language families. In some cases, these patterns do not apply to the Chinese language. This may be fundamentally due to the high-tech dominance of the English language, so LLMs are commonly trained in English. English has incorporated many aspects of European languages (Germanic, French, and Latin) as well as vocabulary from other global languages. Chinese uses a different tokenisation to English, but most LLMs are able to deal with Chinese, perhaps due to plentiful Chinese training data.
(3) We apply Agent-OM-Lite to two biomedical tracks (bio-ml and biodiv). We find that the computation time for these two tracks is significantly longer than for other tracks. This is because the ontologies used in these tracks are large ontologies and Agent-OM always captures syntactic, lexical, and semantic information for each ontology entity. In general, it is a useful practice because it addresses two common matching scenarios: the same concept with different names and different concepts with the same name. However, in some tasks in the biomedical domain, it is rare for different concepts to have the same name, for example, ncit-doid in bio-ml and fish-zooplankton in biodiv. Therefore, it could be worthwhile to initially match only by syntactic matching and to assess intermediate results. For those tasks where performance is excellent, matching could stop there. For those tasks with poor performance, proceeding to the much more computationally-demanding LLM-based lexical and semantic steps could be justified.

3.2. Discussion on ways to improve the proposed system

(1) Agent-OM can be used for both subsumption matching and one-to-many/many-to-many matching. However, such matches are susceptible to the similarity threshold chosen. When similarity is very high, we could declare three matches (i.e. equivalence, subsumption, and inverse subsumption), but we cannot determine which to use. Although one entity can have multiple closely-matched candidates, it is hard to determine the best similarity threshold as a cut-off point. In some cases, one could look for a "gap" in the similarity scores to define the cut-off point, obtaining a different cut-off for each mapping.
(2) Agent-OM is currently using bidirectional validation to reduce LLM hallucinations, but it is not efficient when the input data to the OM system is unbalanced. One of the ontologies may be much larger than another. In such settings, the system should select the smaller ontology as the starting point so that validation can be applied to fewer matching candidates.
(3) After LLM validation, an extra step of human validation could be useful for precise mappings. Although LLMs can serve as oracles acting as domain experts [15], several limitations should be taken into account. For example, the llama-3-8b model may treat "A is the subclass of B" and "B is the subclass of A" as contradictions, even though these two statements indicate the equivalence of A and B.

In the era of LLMs, we believe that there are two pathways to develop modern LLM-based OM systems. One is to explore the new LLM infrastructure, and the other is the LLM fine-tuning. The former often injects external knowledge from retrieval-augmented generation (RAG) [16] into the LLMs, while the latter uses training data to update the internal knowledge of the LLMs. A communication module (e.g.

LLM agent) is the critical component of the LLM infrastructure, while high-quality data is the key to finding the "Aha! moment" for LLM fine-tuning. Recent research focusing on LLM infrastructure includes [1, 17, 18, 19, 20, 21], while LLM fine-tuning is addressed in [3, 22, 23, 24].

3.3. Comments on the OAEI test cases

(1) The namespaces in reference.xml are mixed with "http://knowledgeweb.semanticweb.org/heterogeneity/alignment#>" (with #) and "http://knowledgeweb.semanticweb.org/heterogeneity/alignment" (without #). A script has been provided to normalise the inconsistent use of namespaces according to the Alignment API format [25].
(2) The ontologies used in the OAEI campaign require a cleaning procedure. Some information irrelevant to the OM task needs to be removed, for example, the metadata of the entity (creator and date/time), which is not relevant to the entity's meaning. Including this metadata can confuse the similarity assessment of an entity pair.
(3) A complete reference is the key to ensuring a fair comparison. We identify two primary reasons for the low performance in certain tracks. a) Some reference alignments have missing mappings. We suggest using LLMs as a tool to validate existing correspondences or to discover missing mappings [26]. b) Some ontologies have some entities with properties (e.g. skos:related) that refer to external resources with naming conventions using codes. In this case, the name carries no natural language meaning and may be confusing to LLMs. We suggest removing these references to external ontologies. Alternatively, we could extend Agent-OM to retrieve external ontologies and use them in our matching process.
(4) For machine learning and LLM fine-tuning for OM, data sampling for the training set needs to be diverse with respect to concepts so that LLMs can learn the domain knowledge. For example, if the alignment includes food nutrition, then the training data is expected to include food nutrition concepts. Several examples of data sampling for OM can be found in [27].

3.4. Comments on the OAEI measures

(1) The output formats for OM systems vary. OAEI only accepts the Alignment API format. There is a need to develop a unified pipeline to convert different formats to the Alignment API format.
(2) LLMs are non-deterministic by nature. An update to the current platform may be required to ensure that LLMs are employed in a uniform setting. We suggest introducing a stream "LLM Arena for OM", in which all systems are expected to use the same LLM and hyperparameter settings for the campaign.

Acknowledgments

The authors thank the Ontology Alignment Evaluation Initiative (OAEI) organising committee and track organisers for their help in dataset curation and clarification. The authors thank Jing Jiang from the Australian National University (ANU) for helpful advice on the justification of multilingual tracks. The authors thank the Commonwealth Scientific and Industrial Research Organisation (CSIRO) for supporting this project. Weiqing Wang is the recipient of an Australian Research Council Discovery Early Career Researcher Award (project number DE250100032) funded by the Australian Government.

This is Agent-OM's first participation in the OAEI campaign. According to the OAEI data policy (retrieved October 1, 2025), "OAEI results and datasets, are publicly available, but subject to a use policy similar to the one defined by NIST for TREC. These rules apply to anyone using these data." Please find more details from the official website: https://oaei.ontologymatching.org/doc/oaei-deontology.2.html.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to grammar and spell check, and to improve the text readability. After using the tool, the authors reviewed and edited the content and take full responsibility for the publication's content.

References

[1] Z. Qiang, W. Wang, K. Taylor, Agent-OM: Leveraging LLM agents for ontology matching, Proceedings of the VLDB Endowment 18 (2024) 516–529. doi:10.14778/3712221.3712222.

[2] Z. Qiang, K. Taylor, W. Wang, OM4OV: Leveraging ontology matching for ontology versioning, 2024. URL: https://arxiv.org/abs/2409.20302. arXiv:2409.20302.

[3] Z. Qiang, K. Taylor, W. Wang, J. Jiang, OAEI-LLM: A benchmark dataset for understanding large language model hallucinations in ontology matching, in: Proceedings of the Special Session on Harmonising Generative AI and Semantic Web Technologies co-located with the 23rd International Semantic Web Conference, volume 3953, CEUR-WS.org, Baltimore, Maryland, USA, 2024.

[4] D. Guo, D. Yang, H. Zhang, et al., DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning, Nature 645 (2025) 633–638. doi:10.1038/s41586-025-09422-z.

[5] OpenAI, Open models by OpenAI, 2025. URL: https://openai.com/open-models/.

[6] Z. Qiang, W. Wang, K. Taylor, Precision–Recall–F1 Visualisation, 2025. URL: https://github.com/qzc438/p-r-f1-vis.

[7] OpenAI, gpt-4o, 2024. URL: https://platform.openai.com/docs/models/gpt-4o.

[8] OpenAI, gpt-4o-mini, 2024. URL: https://platform.openai.com/docs/models/gpt-4o-mini.

[9] Meta, llama-3, 2024. URL: https://www.llama.com/models/llama-3/.

[10] OpenAI, ada-002, 2022. URL: https://platform.openai.com/docs/models/text-embedding-ada-002.

[11] G. V. Cormack, C. L. A. Clarke, S. Buettcher, Reciprocal rank fusion outperforms condorcet and individual rank learning methods, in: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, Boston, Massachusetts, USA, 2009, pp. 758–759. doi:10.1145/1571941.1572114.

[12] J. Kim, B. Shin, J. Chung, M. Rhu, The cost of dynamic reasoning: Demystifying AI agents and test-time scaling from an AI infrastructure perspective, 2025. URL: https://arxiv.org/abs/2506.04301. arXiv:2506.04301.

[13] OpenAI, OpenAI rate limits, 2025. URL: https://platform.openai.com/docs/guides/rate-limits.

[14] S. Hertling, J. Portisch, H. Paulheim, MELT - matching evaluation toolkit, in: Semantic Systems. The Power of AI and Knowledge Graphs, volume 11702, Springer, Karlsruhe, Germany, 2019, pp. 231–245. doi:10.1007/978-3-030-33220-4_17.

[15] S. Lushnei, D. Shumskyi, S. Shykula, E. Jimenez-Ruiz, A. d'Avila Garcez, Large language models as oracles for ontology alignment, 2025. URL: https://arxiv.org/abs/2508.08500. arXiv:2508.08500.

[16] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, D. Kiela, Retrieval-augmented generation for knowledge-intensive NLP tasks, in: Proceedings of the 34th Annual Conference on Neural Information Processing Systems, volume 33, Curran Associates, Inc., Vancouver, British Columbia, Canada, 2020, pp. 9459–9474.

[17] S. Hertling, H. Paulheim, OLaLa: Ontology matching with large language models, in: Proceedings of the 12th Knowledge Capture Conference 2023, ACM, Pensacola, Florida, USA, 2023, pp. 131–139. doi:10.1145/3587259.3627571.

[18] H. Babaei Giglou, J. D'Souza, F. Engel, S. Auer, LLMs4OM: Matching ontologies with large language models, in: The Semantic Web: ESWC 2024 Satellite Events, Springer, Hersonissos, Crete, Greece, 2024, pp. 25–35. doi:10.1007/978-3-031-78952-6_3.

[19] S. Zhang, Y. Dong, Y. Zhang, T. R. Payne, J. Zhang, Large language model assisted multi-agent dialogue for ontology alignment, in: Proceedings of the 2024 International Conference on Autonomous Agents and Multiagent Systems, IFAAMAS, Auckland, New Zealand, 2024, pp. 2594–2596.

[20] M. Taboada, D. Martinez, M. Arideh, R. Mosquera, Ontology matching with large language models and prioritized depth-first search, Information Fusion 123 (2025) 103254. doi:10.1016/j.inffus.2025.103254.

[21] L. Nguyen, E. Barcelos, R. French, Y. Wu, KROMA: Ontology matching with knowledge retrieval and large language models, in: The Semantic Web – ISWC 2025, Springer, Nara, Japan, 2025, pp. 629–649. doi:10.1007/978-3-032-09527-5_34.

[22] Y. He, Z. Yuan, J. Chen, I. Horrocks, Language models as hierarchy encoders, in: Advances in

Neural Information Processing Systems, volume 37, Curran Associates, Inc., 2024, pp. 14690–14711. doi:10.52202/079017-0469.

[23] G. Sousa, R. Lima, C. Trojahn, Complex ontology matching with large language model embeddings, 2025. URL: https://arxiv.org/abs/2502.13619. arXiv:2502.13619.

[24] H. Yang, J. Chen, Y. He, Y. Gao, I. Horrocks, Language models as ontology encoders, in: The Semantic Web – ISWC 2025: 24th International Semantic Web Conference, Springer, Nara, Japan, 2025, pp. 443–461. doi:10.1007/978-3-032-09527-5_24.

[25] J. David, J. Euzenat, F. Scharffe, C. Trojahn dos Santos, The Alignment API 4.0, Semantic Web 2 (2011) 3–10. doi:10.3233/SW-2011-0028.

[26] Z. Qiang, K. Taylor, W. Wang, How does a text preprocessing pipeline affect ontology matching?, 2024. URL: https://arxiv.org/abs/2411.03962. arXiv:2411.03962.

[27] S. Hertling, E. Norouzi, H. Sack, OAEI machine learning dataset for online model generation, in: The Semantic Web: ESWC 2024 Satellite Events, Springer, Hersonissos, Crete, Greece, 2024, pp. 239–243. doi:10.1007/978-3-031-78952-6_34.